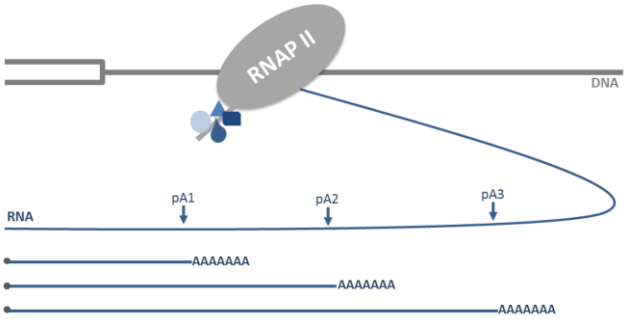


Regulation of the alternative polyadenylation in T cells

Inês do Lago e Baldaia
Mestrado em Biologia Celular e Molecular
Departamento de Biologia
2013
Orientador
Alexandra Moreira, PhD, IBMC-UP

M

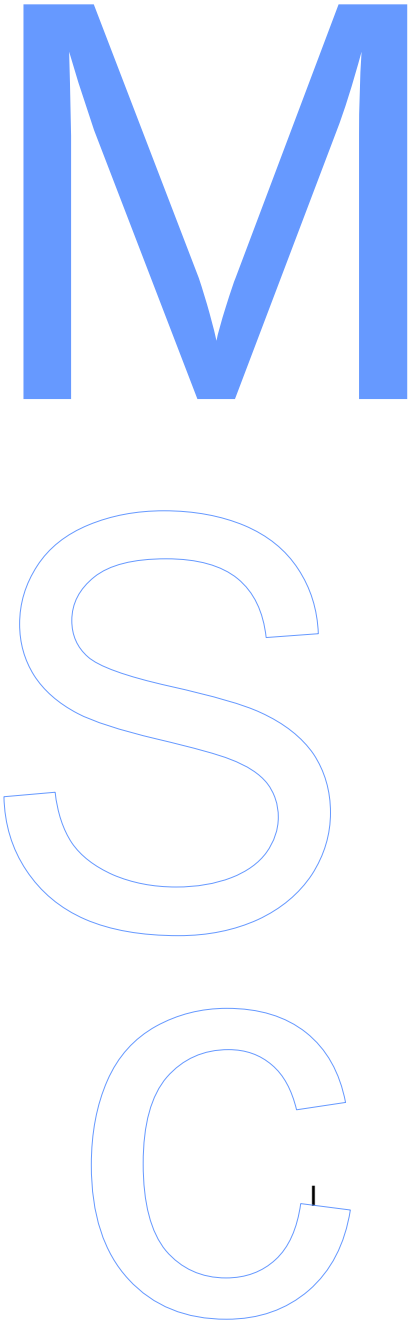
S

C

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



Acknowledgments

O meu primeiro obrigado é para a Alexandra, minha orientadora, por me ter aceitado como sua aluna, e me ter dado a oportunidade de ver um grupo a nascer. Sinto-me privilegiada por ter participado na fundação do Gene Regulation. Obrigada pela paciência e apoio, tanto científico como pessoal, mas acima de tudo, obrigada por me ter mostrado que na ciência é preciso trabalhar arduamente, mas sempre com optimismo e vontade, e me ter ajudado a ultrapassar o pessimismo.

Obrigada à Inês! Fizeste com que o trabalho parecesse tempos livres. Obrigada pelo apoio científico, pela boa disposição, pela amizade… sem ti o IBMC era muito mais chato.

Um especial agradecimento à Isabel (só é pena não teres chegado mais cedo!), Vânia, Rita, Liliana, pela paciência e ajuda científica. Ao resto do GR, Ana, Andrea, Eder, e ao resto do CAGE e Alexandre, por me terem recebido e por todo o apoio e ajuda.

E claro, à minha família, ao Bruno e aos meus amigos, por me ajudarem a manter os pés na terra, optimismo e segurança.

Resumo

O processamento da extremidade 3’ do pré-mRNA é um importante mecanismo de regulação genética, e grande parte dos genes em mamíferos apresenta mais do que um possível local de poliadenilação, produzindo diferentes moléculas de mRNA e contribuindo para a heterogeneidade transcricional. Os locais de poliadenilação alternativos (PAS) podem localizar-se em intrões ou diferentes exões, produzindo diferentes isoformas proteicas, ou serem restritos ao 3’ UTR, alterando o comprimento deste e por conseguinte locais de ligação de factores proteicos ou alvos de microRNAs. A poliadenilação alternativa (APA) é regulada ao nível da cromatina e mRNA, e pode ser influenciada por diferentes condições biológicas, como diferenciação celular, desenvolvimento e cancro.

CD5 é uma glicoproteína transmembranar constitutivamente expressa em linfócitos T e num subgrupo de linfócitos B, B1a. O CD5 tem uma função inibitória na activação do TCR e BCR, e a sua desregulação desencadeia uma resposta imune deficiente.

As extremidades 3’ do CD5 foram mapeadas numa linha tumoral de células T (Jurkat) e foram identificadas três principais isoformas de mRNA, diferindo no comprimento do 3' UTR – pA1, pA2 e pA3. A isoforma mais curta (pA1), produzida por PAS teoricamente mais fracos, é a mais expressa tanto em células Jurkat como células T primárias, enquanto a isoforma mais longa, produzida pelo uso do PAS teoricamente mais forte, e a menos expressa. No entanto, as proporções das diferentes isoformas variam entre os dois tipos de células. A isoforma pA3 apresenta uma maior percentagem em células primárias, enquanto a isoforma pA2 é favorecida em células Jurkat. Curiosamente, esta isoforma apresentou a maior eficiência de tradução num trabalho anterior, nestas mesmas células, sugerindo um mecanismo de optimização de produção de proteína.

A estreita relação entre a tradução e os mecanismos de processamento de mRNA traduz-se nas múltiplas funções das proteínas intervenientes. O papel de factores de *splicing* na APA do CD5 foi investigado através de *knockdown* por siRNA, e foi possível mostrar que a proteína PTB favorece a selecção do sinal pA1 e inibe o uso do sinal pA2, e que SRSF6 favorece o sinal pA1 mas inibe o sinal pA3. Mostrou-se também que as regiões vizinhas dos três PAS têm diferentes padrões de ligação de proteínas. O estado da cromatina também mostrou influenciar a APA do CD5, uma vez que baixos níveis de acetilação aumentam o uso das isoformas pA2 e pA3.

Os nossos resultados sugerem que o CD5 é fortemente regulado pela APA através de vários mecanismos distintos.

Palavras-chave: Poliadenilação alternativa; 3’ UTR; células T; células Jurkat; isoformas de mRNA; Activação de células T; Modificações de histonas; Proteínas que se ligam ao RNA (RBPs).

Abstract

Alternative pre-mRNA 3' end processing plays a major role in gene expression and most mammalian genes have multiple polyadenylation (pA) sites that can be differentially selected, thus contributing to the enrichment of the transcriptional heterogeneity. The alternative pA signals (PAS) can be located in an internal exon or an intron, yielding different protein isoforms, or be positioned in tandem at the 3' UTR, resulting in mRNAs with different 3' UTR lengths, with the inclusion or exclusion of binding sites for specific RNA binding proteins (RBPs) or microRNA target sites. APA is regulated at the chromatin and mRNA levels, and may vary under specific conditions, such as cellular differentiation, development and cancer.

CD5 is a transmembrane glycoprotein with three scavenger receptor cysteine rich (SRCR) domains, constitutively expressed in T-cells and in a subset of B-cells, B1a cells. CD5 has an inhibitory function in the TCR and BCR activation and its deregulation triggers a defective immune response.

We mapped the mRNA 3' ends of CD5 in a Jurkat T cell line and identified three major mRNA isoforms differing in the 3' UTR length – pA1, pA2 and pA3. The shortest isoform (pA1), defined by the predicted weakest PAS, is the most expressed one, both in Jurkat and primary T cells, while the longest isoform (pA3), defined by the canonical PAS, is the least expressed. Even though, the proportions of the different CD5 APA isoforms vary in the two cell types. The pA3 isoform presents a higher percentage in primary T cells, relative to Jurkat, while the pA2 isoform is favored in Jurkat cells. Surprisingly, this is the most efficiently translated isoform, as reported in a previous work, which suggests a protein production optimization mechanism.

The tight relation between the different mRNA processing mechanisms and transcription relies on activity of protein factors that often play multiple functions. We investigated the role of splicing factors in CD5 APA by siRNA knockdown, and we showed that PTB favors the selection of the pA1 signal and inhibits the usage of the pA2 signal, and that SRSF6 also favors the pA1 signal but inhibits the pA3 signal. Moreover, we showed that the surrounding regions of the three PAS have different patterns of binding of RNA binding proteins. The chromatin state also showed to influence CD5 APA, as low levels of acetylation increase the usage of pA2 and pA3 isoforms.

Our results suggest that CD5 is regulated by APA through several distinct mechanisms, and that they may work together in a PAS specific manner.

Keywords: Alternative Polyadenylation; 3' UTR; CD5; T cells; Jurkat cells; mRNA isoforms; T cells activation; Histone modifications; RNA-binding proteins.

Table of Contents

Figures and Tables List

Abbreviations

3’ RACE	Rapid Amplification of cDNA 3’ end
5’ UTR / 3’ UTR	5’ and 3’ Untranslated Regions
aa	Amino Acid Residue
Ab	Antibody
APA	Alternative Polyadenylation
BCR	B cell Membrane Receptor
CD	Cluster of Differentiation
FACS	Fluorescence-Activated Cell Sorting
KD	Knockdown
kDa	Kilo Dalton
miRNA	MicroRNA
mRNA	Mature Messenger RNA
PAS	Polyadenylation Signal
PBMC	Peripheral blood mononuclear cell
PHA	Phytohemagglutinin
Poly(A)/pA	Polyadenylation
RBP	RNA-binding Protein
siRNA	Small Interference RNAs
SRCR	Scavenger Receptor Cysteine Rich
SRSF	Serine/Arginine-Rich Splicing Factor
TCR	T Cell Membrane Receptor
TSA	Trichostatin A
A	Adenine
C	Cytosine
G	Guanine
K	Keto nucleotide (G or T)

M	Amino nucleotide (A or C)
R	Purine
T	Tyrosine
U	Uracil
Y	Pyrimidine

Introduction

The Immune System and Adaptive Response

The immune system refers to all the structures and processes that protect an organism from biological invaders or chemical elements which could cause a disease. It is composed of two different types of defense, one less specific and other more specific, with different functions and timing of response. The less specific system, innate immunity, is the first line of defense against a broad set of pathogens and includes anatomic and physiologic barriers, the inflammatory process and the cellular response through specialized cells performing phagocytosis [1].

The adaptive immune system is the second line of defense and can recognize and eliminate specific foreign molecules and microorganisms – antigens. This system is highly diverse and specific as it can identify billions of antigens. It also distinguishes self from non-self-antigens, preventing autoimmune responses. The other main feature is the immunologic memory in which, after the first encounter with an antigen, memory cells specific for that antigen are maintain in the blood circulation so that in a second encounter the system can produce a faster and stronger response [1].

The main players in the adaptive immune system are the B and T lymphocytes. The B lymphocytes differentiate and maturate within the bone marrow and, when activated, perform the humoral immune response, which consists in the release of antibodies to the bloodstream that neutralize the antigens and activate other immune pathways. T lymphocytes differentiate and maturate in the thymus, where they start to express the T cell membrane receptor (TCR) which recognizes specific antigens linked to major histocompatibility complex (MHC) membrane molecules. There are two major classes of these molecules, class I that is expressed by nearly all nucleated cells and class II that is expressed only by the antigen-presenting cells (macrophages, B lymphocytes and dendritic cells). The T helper (T_H) cells, which express the CD4 membrane glycoprotein, interact with antigen–class II MHC molecule complexes and are activated to secrete different cytokines in order to induce the activity of other cells, such as macrophages and B cells. On the other hand, T cytotoxic (T_C) cells, which express the CD8 membrane glycoprotein, interact with antigen–class I MHC molecule complexes and, in the presence of specific cytokines, differentiate into cytotoxic T lymphocytes (CTL). These have cytotoxic or killing activity against the foreign antigen–class I MHC molecule complexes presenting cells, such as infected or tumor cells. For the correct antigen recognition and T cell activation several accessory receptors are

needed besides the TCR and CD4/CD8, that are involved in the strength of the bond between T cells and the antigen-presenting cells or/and in the signal transduction [1].

The CD5 molecule

CD5 is one of the accessory receptors referred above, with the function of regulating the immune response in T and B lymphocytes. Much attention has been channeled to the study of this molecule in the last 20 years, although little is known about the CD5 expression regulation. This understanding is essential for therapeutic research since this molecule is associated with several pathological processes.

The human CD5 is a type I transmembrane (single pass molecule with the C-terminal directed to the cytoplasm) glycoprotein with a 67 kDa molecular mass [2]. It is a member of the highly conserved scavenger receptor cysteine rich (SRCR) superfamily, containing in its extracellular region three SRCR domains. Its domains are classified as type B, along with the SRCR domains of CD6 and Spα, for example, characterized by the presence of 8 cysteine residues and being encoded by a single exon (in contrast, type A domains contain 6 cysteine residues and are encoded by more than one exon) [3]. The SRCR domains are retained in the natural soluble CD5 form identified in the human serum, proposed to be originated by proteolytic cleavage [4]. The CD5 gene is located in chromosome 11q13 and contains 11 exons and 10 introns (*Ensembl*). After cleavage of the 24 amino acid residues (aa) signal peptide, the extracellular region is composed of 348 aa with the three SRCR domains encoded by exons 3, 5 and 6. The transmembrane domain of 30 aa is encoded by exon 7 and the intracellular region composed by the last 93 aa is encoded by exons 8 to 10 (predicted regions from *NCBI* and *UniProt*).

In humans, CD5 is constitutively expressed in thymocytes and mature T lymphocytes, but is restricted to a subpopulation of B lymphocytes, B1a cells [5, 6]. This subpopulation represents most of the B cells in the fetus and umbilical cord blood, but it decreases with age representing only 10-25% in the peripheral blood and slightly higher percentages in the peritoneal cavity and tonsil [6, 7]. CD5 is also expressed in a recently described regulatory B cells subset in mice, CD1dhiCD5+ B cells [21].

In B-cell, the pan-B cell surface molecule CD72 [8] and the V_H framework regions of immunoglobulins or Fab fragments [9] showed to interact with CD5. Also expressed in activated T cells, a molecule different from CD72 was identified to interact with CD5, CD5L (CD5 ligand) [10]. Other molecules have been shown to interact with the extracellular region of CD5, but it was not yet identified a consensual physiological

relevant CD5 ligand. Moreover, Brown and Lacey showed that CD5 can interact with itself through species specific homophilic binding [11].

The intracellular region is highly conserved between human and mouse [12] and contains specific motifs involved in signal transduction, suggesting an important role for this region that may be associated with CD5 function. Of the four potential tyrosine phosphorylation sites, residues Y429 and Y463 showed to be phosphorylated, probably by the protein tyrosine kinase Lck, after T cell activation [13]. Along with Y441, Y429 is part of an imperfect immunoreceptor tyrosine-based activation motif (ITAM) [13]. The membrane-proximal residue Y378, located at an imperfect immunoreceptor tyrosine-based inhibition motif (ITIM), showed to associate with the protein tyrosine phosphatase SHP-1 and inhibit the TCR/CD3 activity [14]. Serine residues S458, S459 and S461, located at the C-terminal region, are phosphorylation targets of the casein kinase II (CKII), which in turn showed to be associated with inhibition of T cell activity [15, 16]. Other molecules such as c-CBL [17, 18] and Ras GTPase activating protein (RasGAP) [18] also showed to associate with CD5.

CD5 functions

Although was initially thought that CD5 functioned as a positive signal receptor [19, 20], it was later characterized as an inhibitory effector. Accumulating evidence support this function, such as CD5-deficient mice show higher T cell proliferation and hyperresponsive thymocytes to stimulation through the TCR [21, 22], and T cells with low CD5 levels present exaggerated TCR-induced activation [23]. CD5 co-localizes with the CD3/TCR complex in the immunological synapse and showed to decrease the calcium response [14, 24]. Moreover, CD5 stimulation increases Fyn phosphorylation, which in turn inhibits ZAP70, an important factor in the TCR signaling cascade [25].

CD5 is expressed since the earliest thymic progenitors to the mature T lymphocytes [5, 26]. The hematopoietic progenitors that arrive to the thymus with no T-cell specific surface markers (TCR and CD4 or CD8) – double negatives (DN) – express low levels of CD5 in a TCR-independent manner [5]. In the following stages, CD5 expression becomes modulated by the pre-TCR and TCR activity [26]: as DN cells differentiate into double positive (DP) cells (simultaneous expression of CD4 and CD8), CD5 expression increases and is maintained in an intermediate level by the low affinity pre-TCR/MHC interaction [5]; as DP cells became single positives (SP), they undergo positive selection, by which the ones unable to recognize the self–MHC molecules are eliminated, and depending on the MHC recognized, I or II, become

engaged to the CD8 or CD4 lineage; the next step is negative selection, when those that have an abnormal strong reaction to the self-MHC or recognize self-antigens are also eliminated [1, 27]. The inhibitory activity of CD5 appears affect the choice of the negative or positive selected thymocytes. For example, thymocytes that do not bind self-MHC or do so with very low affinity are eliminated in the presence of CD5, instead of being incorrectly positively selected [26]. On the other hand, it was shown that positively selected thymocytes with high affinity TCR/MHC interaction upregulate CD5 and that in its absence, are instead eliminated [26]. Azzam *et al.* proposed that CD5 has a regulatory function in these selection processes in a TCR's clonotypes specific manner, promoting a tendency to sort high affinity TCRs to form active and useful mature T cells. So, although CD5 knockdown mice show no significant changes in thymocyte maturation, different groups of thymocytes undergo a fate shift either favoring positive selection or negative selection [5, 26].

Given CD5 function in the selection of the TCR repertoire during T cell differentiation and the control of the TCR activation threshold in mature T cells, this molecule presents itself as an important regulator of the T cells tolerance, and its deregulation can lead to an exaggerate immune response [23].

Furthermore, the CD5 effect in T cell activity seems to have consequences in the activation-induced cell death (AICD), a homeostatic mechanism that controls the pool of activated T cells and induces T cell tolerance [28]. AICD is the process by which activated T cells overexpress Fas and its ligand (FasL), which activate caspases and pro-apoptotic effectors, leading to cell death [1]. Friedlein *et al.* showed that tumor-reactive circulating T cells expressed higher levels of CD5 than tumor-infiltrating lymphocytes (TIL) and, in fact, the first cells also expressed lower levels of FasL and caspase-8. Therefore, these $CD5^{high}$ lymphocytes were less susceptible to AICD, and this feature could be reversed by incubation with an anti-CD5 mAb [28], but also showed weaker Ca^{2+} response facing tumor cells than TIL [29]. To overcome the tumor evasion, it seems that TIL decrease the CD5 levels to adjust their sensitivity and increase the tumor-specific T cell response, although it renders them more susceptible to AICD [30].

In B1a cells, the immune response mediated by the BCR is also negatively regulated by CD5 [31]. These cells characteristically produce natural IgM with low affinity polyreactivity to auto-antigens, and high levels of the anti-inflammatory cytokine IL-10 [32].

CD5 and human disorders

The B1a cells have been associated with several autoimmune models. Experimental autoimmune encephalitis (EAE) and inflammatory bowel disease are two examples of autoimmune models that showed a protective role of these cells through the production of the anti-inflammatory cytokine IL-10 [29, 33]. In contrast, the B1a cells production of natural polyreactive antibodies for auto-antigens and enhanced antigen-presenting ability are two features that seem to contribute to autoimmune processes [33, 34]. Also, the role of CD5 in the regulation of the TCR activation was associated with a protective function of CD5 in an EAE model [35].

As described above, CD5 is involved in the antitumor immune response, but is also associated with specific neoplastic disorders. The most striking example is the B cell chronic lymphocytic leukaemia (B-CLL), which is characterized by an accumulation of CD5 positive B cells, with increased levels of survival factors and IL-10, that enhances B cell proliferation, stimulated by a CD5 and PKC (phosphorylation of protein kinase C) dependent signaling pathway [29, 36].

3' end processing

For most of the eukaryotic mRNAs to be stable, able to be exported from the nucleus and be correctly translated, the pre-mRNA molecules must undergo a series of co-transcriptional modifications to become mature mRNAs. These modifications are the 5' capping, addition of a 7-methylguanosine (m^7G) to the 5' end of the pre-mRNA; the splicing, removal of introns and ligation of the exons, including the 5' and 3' untranslated regions (5' and 3' UTR); and 3' end processing. In this thesis, we will focus on this last mechanism, also referred as polyadenylation, and how its regulation can influence gene expression.

Polyadenylation is a widespread mechanism in eukaryotes [37, 38], although it does not generally happen in rRNAs (ribosomal RNA), and is less consensual in prokaryotes [39, 40]. It is a co-transcriptional two step mechanism consisting in the endonucleolytic cleavage of the pre-mRNA and polymerization of a poly(A) (adenosine monophosphate) tail [41, 42]. This tail can have between 250 and 300 adenosines in humans, but it may vary in a specie-specific manner, and its size functions as a mark for degradation [43]. The cleavage site in the pre-mRNA, pA site, is defined by several *cis*-elements that function as binding sites for essential cleavage and polyadenylation factors. The pA site region is composed by an A-rich hexamer, the pA signal (PAS), an U/GU-rich downstream sequence element (DSE), and U-rich upstream sequence

elements (USE) [42]. The canonical PAS is the AAUAAA sequence, present in almost 60% of the human pre-mRNAs, but several variants are also used, such as AUUAAA in ~15% and other single nucleotide variants of the hexamer in ~12% of the cases [44]. This hexamer is located 10-30 nt upstream of the pre-mRNA cleavage site and is recognized by the CPSF (cleavage and polyadenylation specific factor), while the DSE is recognized by the CstF (cleavage stimulation factor) [42]. The binding of these two proteins recruits other important factors, as CFI and CII (cleavage factors I and II) and PAP (polyadenylation polymerase), to compose the polyadenylation core machinery [42]. The CTD (carboxy-terminal domain) of the RNA polymerase II (RNAP II) is also important, being required as a recruiter and platform for necessary pA factors [45]. As for the cleavage site, it is characteristically immediately downstream of a YA (pyrimidine-adenine) motif but it can vary in several nucleotides upstream or downstream depending on the exact position of the defining sequences referred above [46].

Alternative Polyadenylation

Every mRNA molecule produced is the result of a pA site selection undertake by the transcription/RNA processing machinery. This process is called alternative polyadenylation (APA). This phenomenon is observed in more than 70% of all mammalian, *Saccharomyces cerevisiae* and *Arabidopsis thaliana*'s genes and about half of the genes of *Drosophila*, Zebrafish and *C. elegans* [47, 48]. Furthermore, many genes present the same pattern, in which the most distal PAS is a stronger consensual PAS, such as the canonical hexamer AAUAAA, and the most proximal is a weaker and less conserved sequence. This distribution suggests that distal PAS function as a default mechanism to ensure the transcription termination and mRNA 3' end processing and therefore the correct maturation of the mRNA [47].

Types of APA

There are two main types of APA: UTR-APA, in which the alternative PAS are located only in the 3' UTR, leading to mRNA isoforms differing on the 3' UTR length without affecting the coding region, and CR-APA (coding region APA), which yields mRNA isoforms with different coding regions [49]. The last type is subdivided in three categories: alternative terminal exon APA, in which alternative splicing removes a PAS altering the last exon; intronic APA, in which a cryptic intronic PAS is used when a

5' splicing site is skipped [47]; and internal exon APA, in which an alternative PAS in located in an internal exon (Fig. 1) [50].

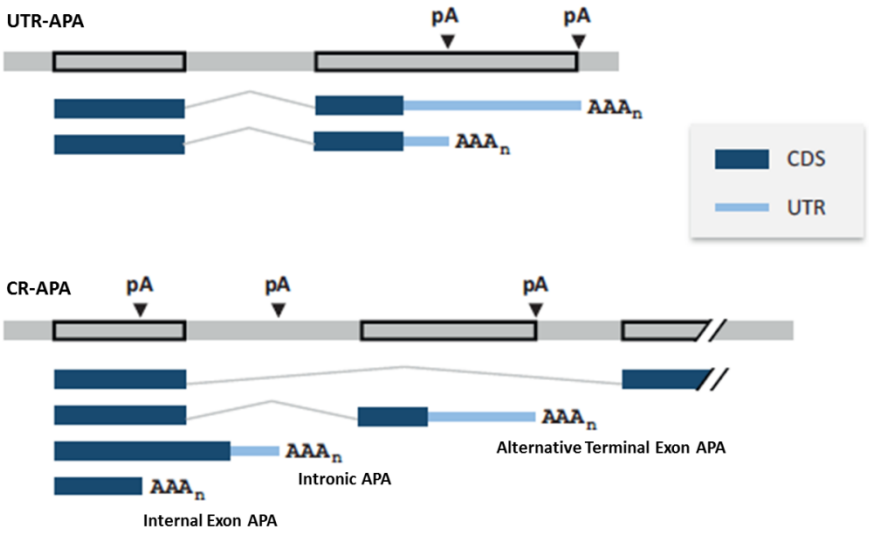


Figure 1. Types of APA mechanisms based on the classification by Elkon *et al*., 2013. UTR-APA refers to the presence of alternative PAS in tandem in the 3' UTR; CR-APA refers to alternative PAS localized in different exons or introns that leads to different coding sequences. Figure adapted from Tian and Manley, 2013.

The CR-APA alters the protein sequence in a qualitative manner, while UTR-APA can affect gene expression through mRNA stability, translation efficiency and localization, by changing the available miRNA target sites and RBPs binding sites in the 3'UTR, that regulate these processes [51].

APA modulation in biological processes

It has been extensively reported that APA is a dynamic mechanism as the alternative mRNA isoforms are differently expressed under different biological conditions. In the last ten years, starting with EST data, several results of genome-wide analysis have been published, showing tendencies in the pattern of the usage of alternative PAS in a tissue-specific manner and in specific biological conditions. For example, transcripts of the nervous system, brain, bone marrow and uterus have a preferential usage of the distal PAS, while in retina, ovary, placenta and blood the most proximal PAS are used [52]. This distribution seems to indicate a preference of the longest mRNA isoforms in tissues associated with cellular differentiation and the shortest in tissues associated with cell proliferation. In fact, Sandberg *et al.* showed a

global shortening of the mRNA isoforms after activation of murine T_H cells associated with the proliferation state and increased gene expression, and further showed that in some cases, full-length 3' UTR displayed less protein expression [53]. Moreover, Jin and Tian showed that induction of dedifferentiation of a somatic cell to a pluripotent stem cell was accompanied by a global 3' UTR shortening, while the contrary showed a global lengthening [54]. An overall concept was thus constructed indicating that proliferation, dedifferentiation and cell transformation states are associated with usage of the proximal PAS, and development and differentiation states are associated with usage of the distal PAS [49].

APA in diseases

There are already described examples of APA involved in human diseases, particularly mutations that lead to a loss or gain of PAS function, or their strengthening or weakening. IPEX (immune dysfunction, polyendocrinopathy, enteropathy, X-linked) is caused by different mutations and deletions of the FOXP3 gene. A mutation in the first PAS (AAUAAA → AAUGAA), which is skipped, increases the usage of the distal PAS, and yields a less stable mRNA isoform that reduces the protein expression of the transcription factor FOXP3 [55].

The proto-oncogene cyclin D1 has two major mRNA isoforms: the longer one results from a 3' UTR PAS, cyclin D1a, and the shorter one results from alternative splicing that retains an intron with intronic PAS, cyclin D1b [56]. The shorter isoform lacks an important phosphorylation site essential for the nuclear exportation, and the protein retention in the nucleus is associated with higher transformation capacity [57]. Also, point mutations showed to create a premature PAS in the cyclin D1a isoform 3' UTR, which was associated with bad prognostic in mantle cell lymphoma (MCL) patients [58]. Similarly to cyclin D1, it has been observed a global 3' UTR shortening, especially in proto-oncogenes, in cancers in comparison to normal tissues, that seems to strongly associate with the proliferation maintenance of the cells [47, 59].

There is also some evidence of the influence of APA in various endocrine diseases [60]. For example, higher usage of the transcription factor 7-like 2 (TCF7L2) proximal PAS seems associated with higher risk of type 2 diabetes.

Regulation of APA

Several levels of regulation contribute to the selection of the PAS to use by the cellular machinery. The *cis*-elements composing the PAS define a strong or weak

signal, the protein factors, *trans*-elements, influence its usage, and chromatin features modulate RNAP II transcription and "reading" of the region comprising the PAS (Fig. 2).

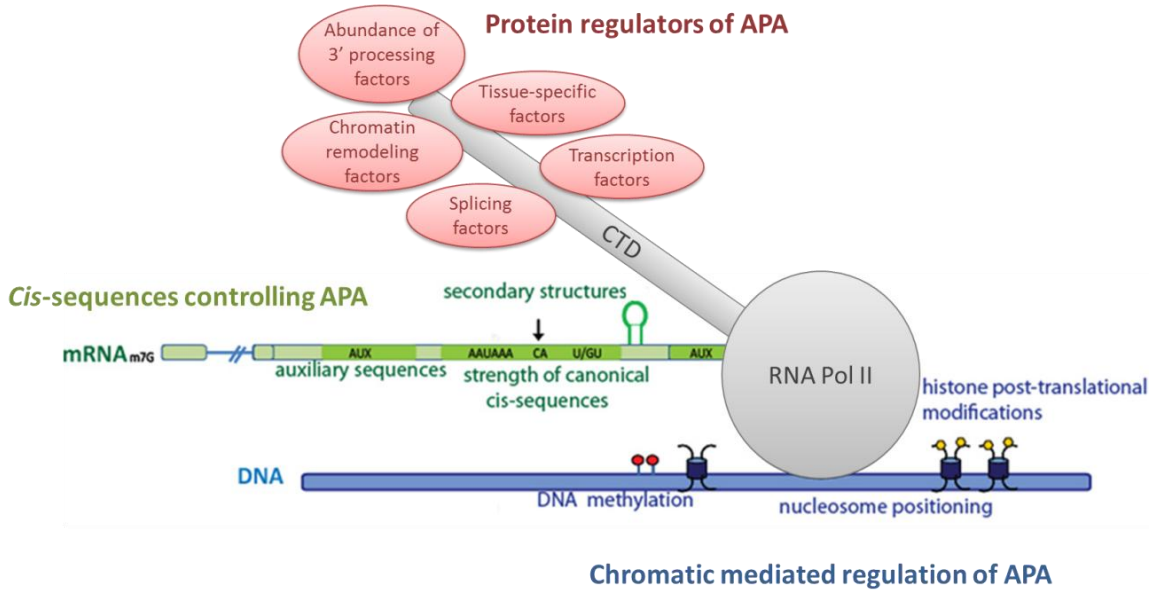


Figure 2. Levels of APA regulation comprise chromatin features, the *cis*-sequences and protein factors. Figure adapted from Di Giammartino *et al*, 2011.

Cis-*elements*

As mentioned above, the sequences that define the pre-mRNA cleavage site include the PAS hexamer and regions upstream and downstream of the signal. The canonical sequence AAUAAA is the most efficiently processed and the most frequently used [44, 61], but the PAS "strength", or efficiency, is also dependent on the surrounding regions [61]. For example, Nunes *et al*. showed that mutations in the DSE of the melanocortin 4 receptor (MC4R) lead to much stronger effects on the 3' end processing efficiency, than the mutation of the respective PAS [62]. Moreover, less efficiently processed PAS variants are frequently surrounded by conserved U- and UG-rich regions, promoting the binding of 3' end processing factors or other *trans*-elements [49]. In APA, the usage of the non-canonical PAS is more frequent than in single PAS genes [44, 62], and the proximal PAS are usually non-canonical, while the distal pA sites use canonical signals [44]. The use of these supposedly weak PAS may facilitate their regulation in APA [44].

Trans-*elements*

Since the pre-mRNA processing occurs co-transcriptionally, the *trans*-elements that influences APA are not only the 3' end processing factors, but also protein factors involved in transcription and in the other pre-mRNA processing mechanisms (see *APA and Transcription* and *APA and Splicing*).

As for the core polyadenylation factors, their upregulation and downregulation was observed in the global 3' UTR shortening described in proliferating cells and the lengthening in differentiating cells [54, 63]. Limiting levels of these factors may favor the usage of stronger PAS, while the presence of higher levels can allow the usage of a weaker PAS [49]. One of the most revisited examples is the effect of the CstF-64 levels, a CstF subunit, in the IgM APA regulation [64]. A secreted isoform of IgM is produced by usage of an intronic PAS and Takagaki *et al.* showed that an increase of CstF-64 concentrations in activated B cells favor the use of the weaker proximal intronic PAS. It is now established that IgM pre-mRNA processing is a complex process, depending not only on CstF levels, but also in competition between the alternative splicing and polyadenylation mechanisms [65]. It was also shown that a set of genes coding for 3' end processing factors contain in their promoters the E2F transcription factor binding site, and that decreased levels of this factor reduced the cleavage at the proximal PAS [66]. These and other studies suggested that overexpression of the 3' end processing machinery leads to higher usage of the weaker, proximal PAS, namely during proliferation. Despite it, it seems more correct to take factor-specific conclusions, since there are studies that show opposite effects for specific factors, as for example the CFIm68 subunit [67].

Chromatin modulation

APA is also regulated at the DNA level, particularly by chromatin structure and epigenetic marks [50], similarly to the transcription and splicing processes. Spies et al. identified a strong nucleosome depletion at the PAS and an enrichment downstream, and that these two features were stronger in highly used PAS [68]. This data raised the hypothesis that the nucleosome enrichment downstream of the PAS had a role in slowing down the RNAP II in that region, promoting its higher usage [69]. The PAS nucleosome depletion was confirmed in a later study that associated this observation with proximal (in case of APA) and unique PAS [70]. Nevertheless, it is still unknown if this depletion is due to the characteristic sequences of the PAS (adenines and thymines), which show lower intrinsic nucleosome affinity, or if it is a consequence of another yet unidentified factor, such as a DNA-binding protein with a

nucleosome-excluding function [49]. A role for epigenetic marks in APA has been suggested, since some examples of genetic imprinting modulating APA were described, such as the *Herc3-Nap1l5* case [71]. *Nap1l5* is a retrogene located in an intron of the host gene *Herc3* that is expressed in the paternal allele, while in the maternal allele its promoter is methylated and thus silenced. In the paternal allele the host gene uses a PAS upstream of *Nap1l5,* whilst in the maternal allele it uses a downstream PAS. Changes in the methylation state of the *Nap1l5* promoter can alter the usage of the two alternative types of PAS of *Herc3*, which further indicates a coupling between RNAP II transcription and APA.

Beside particular examples, global epigenetic marks are also associated with PAS regions, such as the increase of the transcriptional repression marks H3K9me3 and H3R2me2 marks around the PAS [69]. Also, the activation mark H3K36me3, which seems correlated with slower RNAP II elongation rate [72], was shown to be decreased downstream of the PAS, even strongly in PAS that are highly used [69]. These data supports a role for chromatin structure in APA regulation.

APA and Transcription

Although it has been proven that transcription termination is independent of the mRNA 3' end processing, it is well known that both processes are intimately connected [41, 73]. In fact, RNAP II transcription termination depends on functional PAS [74]. Several 3' end processing factors bind the RNAP II CTD [75], such as CPSF, which is recruited since the beginning of transcription by the transcription factor TFIID to RNAP II CTD at the promoter [76], and the yeast Pcf11p subunit of CF IA [77]. The CTD functions as a platform and recruiter, and therefore, is considered a member of the 3' end processing machinery [75, 78]. But the interaction between transcription and 3' end processing is not restricted to the CTD. As Nagaike *et al*. demonstrated, transcription elongation factors, such as PAF1c (RNAP II-associated factor), can enhance the mRNA 3' end processing, and its upregulation causes a usage switch to the proximal PAS [79].

Other important feature of transcription that influences APA is the RNAP II elongation rate, as shown by Pinto *et al*. using a *Drosophila* strain with a point mutation in the larger subunit of RNAP II. This mutation renders the RNAP II 50% slower and promotes the usage of the proximal PAS in six genes, including *polo* [80]. Moreira proposed that the deceleration of RNAP II allows the proximal PAS to be longer exposed to the 3' end processing machinery facilitating its usage rather than the distal

PAS [81]. This model is consistent with the epigenetic data mentioned above, which shows that chromatin marks that slow RNAP II are present near the PAS, particularly in high usage PAS.

APA and Splicing

Similar to transcription, the splicing mechanism also influences APA. It is easy to correlate these two mechanisms since splicing is directly involved in two types of CR-APA, alternative terminal exon APA and intronic APA (see *Types of APA*), by excluding or exposing an exonic or intronic PAS.

Components of the spliceosome, such as U1 and U2 small nuclear ribonucleoproteins (snRNP) and U2 auxiliary factor (U2AF), have individually shown a role in APA. In knockdown studies of U1, which recognizes the 5' splicing site of the intron, it was observed that, not only it repressed splicing, but also promoted premature polyadenylation in cryptic intronic PAS [82]. Interestingly, the consequent polyadenylation was splicing-independent, as the knockdown of other spliceosome components did not show the same effect. Moreover, it is known that U1 can interact with and inhibit PAP [83], and that also binds to the COX-2 proximal PAS USE [84]. A recent wide-genome analysis indicates that U1 plays an important role in APA, as it showed that this molecule protects proximal PAS and its decrease is associated with mRNA 3' UTR shortening [85]. U2 also interacts with PAP, while the auxiliary factor U2AF interacts with CPSF and CFI, and both showed to promote a correct 3' end processing [86].

3' end processing, APA and RBPs

The mRNA processing is also regulated by RBPs and several of them are involved in splicing have shown to be also implicated in APA. One example is NOVA2, a neuron-specific splicing factor RBP, with already identified binding sites in 3' UTRs. When its site is located within 30 nt of the PAS, NOVA2 inhibits its recognition by the 3' end processing machinery, while when it is located in distal elements it promotes the PAS usage [87]. A triplet-repeat expansion in the polyadenylation binding protein nuclear I (*PABPNI*) gene causes autosomal-dominant oculopharyngeal muscular dystrophy (OPMD) [88] and it seems to relate to APA deregulation, since mutations or knockdown of PABPN1 induces a global shift to proximal PAS usage [89]. Other example is the *Drosophila* embryonic-lethal abnormal visual system (ELAV) protein,

which was shown to directly bind and repress the proximal PAS, therefore promoting the distal PAS usage in neuronal tissues [90].

The hnRNPs (heterogeneous nuclear ribonucleoproteins) are proteins that form RNA-protein complexes with the pre-mRNA molecules in the nucleus, and can assist in the correct mRNA processing, nuclear export and metabolism. Several hnRNPs have been associated with 3' end processing regulation (review in [86]), such as hnRNP F, L and H. A particularly important hnRNP in this work is the hnRNP I, also known as PTB (polypyrimidine tract-binding protein). This protein was first reported to bind to the USE element of the *C2* (complement 2) gene by Moreira *et al.*, and be required for efficient pA site cleavage [91]. It was also shown correct levels of PTB in the cells were determinant for 3' end processing, as PTB knockdown led to decrease in pA site cleavage, but it was also shown that overexpression of PTB induced a similar outcome followed by reduced levels of mRNA [92]. Castelo-Branco *et al.* thus proposed a specific model for PTB in the *C2* polyadenylation: its positive function may be due to recruitment of essential 3' end processing factors when PTB levels are balanced; high PTB levels cause a negative effect in polyadenylation, as PTB competes with CstF that, in the case of the *C2* pre-mRNA that lacks a DSE, also binds to the USE. Hall-Pogar *et al*. and Danckwardt *et al.* showed that PTB also binds to the USE of COX-2 and prothrombin encoding genes, respectively, playing a role in the correct 3' end processing [84, 93]. The PTB 3' end processing stimulating mechanisms were further clarified when Millevoi *et al.* showed that PTB facilitates the hnRNP H binding to G-rich auxiliary elements upstream and downstream of the PAS which in turn stimulates the assembly of 3' end processing factors (PAP and CstF), promoting both cleavage and polyadenylation [94]. PTB acts as an auxiliary factor for cleavage/polyadenylation, but was also recently associated with APA of the beta-adducin encoding gene [95]. This gene undergoes APA in a tissue-specific manner and PTB was shown to bind the DSE of the distal PAS, which when deleted induces a shift to higher usage of the cryptic PAS.

Other important family of RNA-binding proteins with functions in splicing is the serine/arginine-rich (SR) protein family, which are currently identified as SRSFs (serine/arginine-rich splicing factors). Some members of this protein family have already been associated also with APA. SRSF1, 2, 5 and 6 (ASF/SF2, SC35, SRp40 and SRp55, respectively) were shown to modulate PAS selection independently of alternative splicing in the glial fibrillary acidic protein (GFAP) APA [96]. Also, in a Rous sarcoma virus (RSV) model, Srp20 and ASF/SF2 showed to promote polyadenylation possibly through interaction with a specific CFI domain [97, 98], and in a later study, Srp20 was associated with transcription termination [99].

Several more RBPs have been identified as APA regulators, operating through a myriad of different interactions and mechanisms. For example, they can compete with 3' end processing factors for binding sites, promote or avoid their recruitment, redirect them to alternative PAS, alter the spatial redistribution or favor post-transcriptional modifications that alter the factors activity [86]. Furthermore, RBPs interact with miRNAs in a bilateral regulation relation, as miRNAs can regulate RBPs expression and vice-versa, and can present a specific interconnection or have a wide-spread effect [100, 101].

Aims

The main goal of this thesis is to gain knowledge about the factors that modulate and control the usage of different PAS, and its consequences in gene expression. CD5 has been increasingly studied in lymphoproliferative and autoimmune disorders contexts, and its importance in these processes has been successively supported. Nevertheless, insight about its expression regulation is still lacking, which could be essential for the understanding of the pathophysiology of such diseases. It was known that CD5 contained at least two alternative PAS in the 3'UTR, and thus it is a good candidate to explore APA. Therefore, we proposed to:

- Map the 3' ends of the CD5 mRNA isoforms to confirm the already reported PAS and/or identify new ones, in T cells;
- Compare the relative use of each PAS in different cellular contexts, as primary T cells *versus* a cancer cell line and resting *versus* activated T cells;
- Investigate the effect of different chromatin states in the CD5 APA and correlate it with PAS usage;
- Identify protein factors, in particular among those with functions in splicing, which influence the CD5 APA, supporting the close involvement of the two processes.

With these goals we hope to shed some light on the molecular mechanisms behind CD5 APA regulation.

Material and Methods

In silico *analysis*

Primers were designed using the Primer-BLAST platform of NCBI, and further analyzed in the OligoAnalyzer 3.1 of Integrated DNA Technologies (eu.idtdna.com). Predicted RNA-binding proteins binding sites were searched using the SFmap predictor. Sequence alignments and pairwise identity percentages were obtained using the default settings of MUSCLE 3.6 running in Geneious v4.8.

Jurkat E6.1 cell culture

E6.1 cells were grown and maintained in 1X Roswell Park Memorial Institute (RPMI) 1640 Medium, GlutaMAX™, supplemented with 10% heat-inactivated fetal bovine serum (FBS), 100 U/mL penicillin and 100 µg/mL streptomycin (all reagents from Gibco). They were incubated at 37° C with 5% CO_2, in T75 culture flasks and passed at a 1:6 ratio, every 2-3 days, maintaining a 0.5-2.5x10^6 cells/mL concentration.

T cell activation

Twenty four hours after E6.1 cells had been passed, they were activated by incubation with PHA (phytohemagglutinin) (Sigma-Aldrich) for 48h at a 10 µg/mL concentration.

Primary T cells extraction

First, we extracted PBMCs (Peripheral Blood Mononuclear Cells), including monocytes, macrophages and lymphocytes, with the Lympholyte®-H reagent (Cedarlane Laboratories). We used defibrinated blood from healthy donor, kindly provided by the Blood Collection Center of Hospital de S. João, Porto. The extraction was performed following the manufacturer's protocol, with the layering of the blood over the Lympholyte®-H reagent, and without platelets removal.

From the PBMCs sample, the T lymphocytes were purified, using the EasySep™ Human $CD4^+$ T cell Enrichment Kit (StemCell), following the Manual EasySep™ protocol using the purple Easysep™ magnet. The obtained cells were diluted in 5 mL of 1X PBS (Phosphate buffered saline).

TSA treatment and RNA fractionation

After extraction of the primary T lymphocytes, they were incubated with TSA (Trichostatin A) for 1h at a 0.5 µM concentration. Cells were collected, washed twice with 1X PBS and resuspended in 1 mL of RSB buffer (10 mM Tris pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$). After 3 min incubation on ice, cells were centrifuged at 4000 rpm for 3 min at 4 ºC, and the supernatant was discarded. The pellet was resuspended in 150 µL of RSBG40 buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 10% glycerol, 0.5% NP-40, 0.5 mM DTT (Invitrogen) and 100U/mL RiboLock RNase Inhibitor (Thermo Scientific)). Cells were centrifuged at 7000 rpm for 3 min at 4 ºC. We analyzed the RNA nuclear fraction, since the short TSA incubation time only causes an effect in the recently transcribed mRNAs. Therefore, the supernatant (cytoplasmatic fraction) was discarded and the nuclei pellet was used to RNA extraction. In every experiment, non-treated T cells, i.e. resting cells, were submitted to the same incubation time with DMSO (solvent of TSA) and fractionation protocol.

Cell counts

A Countess® Automated Cell Counter (Invitrogen) was used to count cells, with 10 µL of a 1:1 mixture of cell suspension and 0.4% Trypan blue (Invitrogen).

RNA extraction and quantification

E6.1 cells RNA was extracted with the PureLink™ RNA Mini Kit (Ambion), following the manufacturer's protocol, and with DNase treatment performed by the On-column PureLink™ DNase Treatment Protocol. The 80 µL DNase mixture contained 0.25 U/µL DNase I recombinant, 1X Incubation buffer (both reagents from Roche) and RNase-free water. RNA was diluted in 30-50 µL of RNase-free water.

RNA was quantified by NanoDrop™ 1000 Spectrophotometer (Thermo Scientific), using RNase-free water as blank.

qRT(Reverse Transcription)-PCR

cDNA synthesis

cDNA was synthesized in a reaction containing 0.4 to 1 µg of RNA, 5 U/µL SuperScript® III Reverse Transcriptase (Invitrogen), 1 U/µL RiboLock RNase Inhibitor (Thermo Scientific), 1X First-Strand Buffer (Invitrogen), 5 mM DTT (Invitrogen), 2.5 M Random Hexamers (5'-NNNNNN-3'; Sigma-Aldrich), 0.5 mM dNTPs mix (Thermo Scientific) and RNase-free water up to 20 µL. First, the water, RNA, Random

Hexamers and dNTPs were mixed and incubated 5 min at 65 ºC to denature the RNA and primers, followed by cooling in ice for 5 min. Then, the rest of the Freagents were added and the reaction was incubated 5 min at 25 ºC to extend primers, 60 min at 50 ºC to extend cDNA strands and 15 min at 70 ºC to inactivate the reaction, in a 48-well TPersonal Thermocycler (Biometra).

qPCR (Real-time PCR)

All primer pairs used were first optimized. The optimization consists in performing qPCR reactions as described below, using a series of dilutions (1:10) of a cDNA template, to construct a standard curve. In a perfect reaction, in each cycle (n) the DNA doubles, so the dilution factor will equal 2^n. For an ideal efficiency the slope of a standard curve generated with 10-fold dilutions would be -3.32. The slope obtained is used to calculate the primer pair efficiency (E) with Formula 1 and the correspondent efficiency percentage (% E) with Formula 2 (Real-Time PCR Applications Guide, Bio-Rad). Different annealing temperatures and primers concentrations were tested in order to obtain the best primer efficiency between 1.9 and 2.1 (90-110%).

Formula 1 $$E = 10^{(-1/slope)}$$

Formula 2 $$\% E = (E - 1) \times 100$$

The quantification of the CD5 isoforms (see *CD5 mRNA isoforms expression* in Results) and the knockdown tests of the siRNA assays (all primers in Table 3) were performed by qPCR. Each reaction contained 1 µL of the synthesized cDNA, 1X iQ™ SYBR® Green Supermix (Bio-Rad) and distilled deionized water (ddH_2O) up to 20 µL, with variable concentrations of primers:

CD5 mRNA isoforms quantification				**Knockdown tests**	
total pA	*pA2+3*	*pA3*	*18S*	*Specific splicing factor primers*	*18S*
0.25 µM	0.25 µM	0.15 µM	0.25 µM	0.125 µM	0.125 µM

Table 1. Primers concentrations for the different qPCR reactions. Forward and reverse primers with the same shown concentration.

Reactions were incubated in the StepOne Real-time PCR System (Applied Biosystems) thermocycler with the following program: 3 min at 95 ºC; 40 cycles of

20 sec at 95 °C and 20 sec at the respective annealing temperature; and 81 cycles increasing 0.5 ºC at each 10 sec starting at 55 ºC.

The endogenous control chosen was the housekeeping gene 18S based on the literature and considering that was the one showing the more consistent amplification (comparing with GAPDH and β-actin) [102, 103]. The results were analyzed using the "ΔC_t method" and "$\Delta\Delta C_t$ method", assuming the maximum efficiency (E = 2), since the results do not suffer significant changes when using the real efficiencies. The "ΔC_t method" generates the expression of the target gene (TG) relative to the endogenous control (reference gene – RG) (Formulae 3a and 3b). These expression levels were also used to generate ratios between the longer isoforms and the total mRNA levels [80]. The "$\Delta\Delta C_t$ method" additionally normalizes the expression level for a specific sample (SS) relatively to a reference sample (RS) (non-treated or scramble transfected cells) (Formula 4 when considering E = 2). This method generates a normalized fold expression, in which values > 1 translate higher expression of the target gene in the specific sample relatively to the reference sample, while values < 1 translate lower expression.

Formula 3a $$\Delta Ct = Ct_{TG} - Ct_{RG}$$

Formula 3b $$\text{Relative Expression} = 2^{-\Delta Ct}$$

Formula 4 $$\text{Normalized Fold Expression} = 2^{-[\Delta Ct_{SS} - \Delta Ct_{RS}]} = 2^{-\Delta\Delta Ct}$$

3' RACE (Rapid Amplification of cDNA Ends)

cDNA synthesis

To map the 3' ends of CD5 mRNA isoforms, cDNA was synthesized from 1 µg of DNase treated RNA extracted from E6.1 cells. The reaction contained 5 U/µL SuperScript® III Reverse Transcriptase (Invitrogen), 1 U/µL RiboLock RNase Inhibitor (Thermo Scientific), 2 µL of First-Strand Buffer 5X (Invitrogen), 3 mM DTT (Invitrogen), 0.6 mM dNTPs mix (Thermo Scientific), and 0.6 µM Poly(dT) adaptor primer and RNase-free water up to 16.5 µL. The RNA, RNase-free water, dNTPs mix and poly(dT) adaptor primer were incubated for 5 min at 65 ºC and 5 min on ice. The rest of the components were added and the total reaction was incubated 55 min at 50 ºC and 15 min at 75 ºC.

First PCR reaction

To 2 µL of the cDNA obtained with poly(dT) adaptor primer, it was added 1X Phusion HF Buffer, 0.012 U/µL Phusion DNA Polymerase (Thermo Scientific), 0.1 µM RACE Fw1 primer, 0.1 µM Universal anchor primer, 0.2 mM dNTPs mix (Thermo Scientific) and ddH_2O up to 50 µL. The thermocycler program was: 1 min at 98 ºC; 35 cycles of 15 sec at 98 ºC, 15 sec at 62 ºC and 1 min at 72 ºC; 7 min at 72 ºC. The reaction product was analyzed by 1.5% agarose gel electrophoresis stained with SybrSafe DNA gel stain (Invitrogen) and using the GeneRuler DNA Ladder Mix (Thermo Scientific).

Nested PCR reaction

From 2 µL of the previous reaction product, a nested PCR was performed using the same protocol and program as the first reaction, but using the RACE Fw2 primer. The product was also analyzed by agarose gel electrophoresis, as described above.

3' RACE using the SMARTer™ RACE kit

The SMARTer™ RACE cDNA Amplification kit (Clontech) was used to synthesized cDNA, using SMARTScribe™ Reverse Transcriptase, and perform the first and nested 3' RACE reactions, using the Phusion DNA Polymerase (Thermo Scientific), following the manufacturer's protocols. The forward primers, RACE kit Fw1 and RACE kit Fw2 (Table 3), were designed according to kit protocol's specificities.

mRNA isoforms sequencing

Cloning of the 3' RACE product

Firstly, 15 µL of the PCR product from the nested 3' RACE reaction was incubated with 10 pmol of dATP (Thermo Scientific) and 1 U of GoTaq® DNA Polymerase (Promega), for 10 min at 72 ºC. Secondly, this product was purified using the QIAquick PCR Purification Kit (Qiagen), following the manufacturer's protocol "PCR purification using a microcentrifuge". Finally, the purified PCR product with addition of dATP was then cloned into the pCR®2.1-TOPO® vector using reagents from the TOPO® TA Cloning® Kit. To 5 µL of DNA, 0.5 µL of vector and 0.5 µL of Salt Solution were added and incubated 20 min at room temperature.

Transformation of competent TOP10 E. coli cells

Heat shock competent TOP10 E. coli cells were transformed with the prepared plasmid DNA, in a 10:1 ratio (5 µL of plasmid DNA to 50 µL of bacteria). This mixture was incubated on ice for 10 min followed by heat shock at 42° C for 1 min and 10 min on ice. After adding 4X LB (Lysogeny broth) medium, the mixture was incubated at 37° C, allowing the cells to start their growth and stabilize. After ~30 min, cells were centrifuged at 3.000 rpm for 2 min. Most of the supernatant was discarded and cells were carefully resuspended in the remaining liquid. The suspension was plated in 50 µg/ml kanamycin LB plates with 20 µL of X-Gal 50 mg/mL and incubated at 37 ºC in a shaker incubator, overnight.

Colony PCR and Plasmid DNA extraction

From de previous plates, white positive and blue negative colonies were submitted to colony PCR. The reaction consisted in 0.3 U of GoTaq® DNA Polymerase, 1X Green GoTaq® Reaction Buffer (Promega), 5 mM $MgCl_2$ (Promega), 1 µM M13 forward and reverse primers (Table 3), 0.4 mM dNTPs mix and ddH_2O up to 10 µL. With a sterile toothpick, the selected colony is dipped into the PCR tube and subsequently used to inoculate new kanamycin LB plates, incubated 37 ºC overnight. From the positive colonies, the ones with different sizes are selected and inoculated in liquid LB medium with 500 µg/ml kanamycin.

Plasmid DNA was extracted from the suspension cultures using the PureLink® Quick Plasmid Miniprep Kit (Invitrogen), following the manufacturer's protocol, and samples were sequenced outside the lab.

Flow cytometry

To quantify surface protein levels of CD69 and CD5 in the E6.1 cell line, cells were collected, washed with 1X PBS and resuspended in 1X PBS containing 0.2% BSA and 0.1% Sodium Azide (PBS/BSA/azide buffer). For each condition and antibody (Ab), $1x10^6$ cells/well were incubated with the respective Ab, following manufacturers' instructions, in 96-well plates, for 25 min on ice. The anti-CD69 Ab, and respective isotype control, were primary labeled, so after the incubation cells were washed twice and resuspended in 400 µL of the PBS/BSA/azide buffer in FACS tubes. The anti-CD5 Ab, and respective isotypes controls, were unlabeled, so after the first incubation, cells were washed twice with the PBS/BSA/azide buffer, followed by incubation with the secondary labeled Ab, for 25 min on ice. After the second incubation, cells were

washed twice and resuspended in 400 µL of the PBS/BSA/azide buffer. Cells were filtered and analyzed in a FACSCalibur (Becton Dickinson).

Antibodies

The antibodies used in the FACS experiments were: Anti-Human CD5 (Y2/178), Mouse Monoclonal IgG1 (Santa Cruz Biotechnology); Polyclonal Rabbit Anti-Mouse Immunoglobulins/RPE Rabbit F(ab')2 (Dako); Mouse IgG1 K Isotype Control (BD Pharmingen); Anti-Human CD69 APC, Mouse Monoclonal IgG1 K (eBioscince); Mouse IgG1 K Isotype Control APC (eBioscince).

E6.1 cells siRNA transfection

Two siRNAs (small interference RNAs) for each target and the scramble sequence (GenePharma), as negative control, from Sigma-Aldrich (Table 2), were transfected into $1x10^6$ E6.1 cells per condition. The Amaxa® Cell Line Nucleofector® Kit V and the Nucleofector® Device, with the X-005 program, were used, following the manufacturer's protocol. siRNAs concentration and time of incubation were optimized for each knockdown target (Table 2).

	siRNA 1 (sense sequence)	siRNA 2 (sense sequence)	siRNA concentration	Incubation time
PTB	GCACAGUGUUGAAGAUCAU[dT][dT]	AACUUCCAUCAUUCCAGAGAA[dT][dT]	50 nM	48 h
SRSF3	CGAAGUGUGUGGGUUGCUA[dT][dT]	GAGAAGUGGUGUACAGGAA[dT][dT]	150 nM	48 h
SRSF6	GUUACUCUUCCAUGUUUAU[dT][dT]	GAACAUCUGGCAGAGACAA[dT][dT]	150 nM	48 h
Scramble	UUCUCCGAACGUGUCACGU[dT][dT]		-	48 h

Table 2. siRNAs sense sequences (5' – 3'), concentration (nM) and time of incubation (h), for each KD target, for Jurkat cells transfection with the Amaxa® Cell Line Nucleofector® Kit V. The concentration used for the scramble is equal to the maximum concentration of the siRNAs used simultaneously.

Nuclear protein extracts

E6.1 cells nuclear protein extracts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents, following the manufacturer's protocol. The extracts were separated in a 12% SDS-PAGE gel, to confirm its integrity.

UV cross-linking assays

In vitro *transcription templates*

Using E6.1 cDNA, three sequences around each CD5 PAS, were amplified using the GoTaq® DNA Polymerase and the 5X Colorless GoTaq® Reaction Buffer (Promega), in a 150 µL PCR reaction, following the manufacturer's protocol. Each forward primer contained the T7 promoter at its 5' end. PCR products were separated in a 1% agarose gel to confirm the amplification of only one band, correspondent to the selected sequence. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), following the manufacturer's protocol "PCR purification using a microcentrifuge".

In vitro *transcription of radiolabeled RNA*

The radiolabeled RNA sequences were transcribed in reactions with 500 ng of the template DNA, 0.5x Transcription Optimized Buffer (Promega), 16 U of T7 RNA polymerase (Promega), 10 mM DTT (Promega), 0.4 mM rATP and rGTP, 12 µM rCTP, 0.08 mM rUTP (Thermo Scientific), 20 µCi [α-32P] CTP, 20 µCi [α-32P] UTP (PerkinElmer) and RNse-free water up to 12.5 µL. The reactions were incubated 2 h at 37ºC. The radiolabeled probes were purified by illustra ProbeQuant G-50 Micro Columns (GE Healthcare), following the manufacturer's protocol.

UV cross-linking

Each RNA probe, 1000 cpm (counts per minute), was incubated with the 30 µL of prepared E6.1 cells nuclear extracts, in a reaction containing 30 mM HEPES pH 7.9, 1.56 mM $MgCl_2$, 0.5 mM ATP, 20 mM Creatine Phosphate and 2.5% polyvinyl alcohol (PVA). Samples were incubated 20 min at 30 ºC and UV cross-linked in Hoefer UVC 500 UV Crosslinker (254 nm, 9 min, 4 cm from light source) on ice. Five µL of Ambion® RNase Cocktail™, containing 2.5 U of RNase A and 100 U of RNase T1, were added to each reaction, and incubated at 37ºC for 30 min.

Immunoprecipitation (IP) of PTB

Thirty µL of the previous UV cross-linked reactions (inputs) were incubated with 5 µL of the anti-Human PTB rabbit polyclonal antibody, kindly provided by Chris Smith's lab, 35 µL of Protein A-Sepharose® 4B beads (Fast Flow from *Staphylococcus aureus*; Sigma-Aldrich) and 130 µL of Buffer D (20 µM HEPES pH 7.9, 0.2 mM EDTA, 20% glycerol, 0.05% NP-40, 0.1 M KCl), at 4 ºC, overnight. Samples were centrifuged

at 2000 rpm for 10 sec and the supernatant was discarded, and the pellet was washed three times with Binding Buffer I (20 µM HEPES pH 7.9, 150 µM NaCl and 0.05% Triton X-100) and II (20 µM HEPES pH 7.9, 150 µM NaCl and 0.4% Triton-X-100).

Five µL of the inputs and the IP reaction were loaded on a 12% SDS-PAGE. Gels were fixed for 20 min at room temperature in a fixing solution, dried and exposed to a high performance chemiluminescence film (GE Healthcare).

Name	Sequence (5' – 3')
Poly(dT) adaptor primer	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTVN
Universal anchor primer	GACCACGCGTATCGATGTCGAC
RACE Fw 1	CATGGGGCTCAGAGGCTGTA
RACE Fw 2	AACTGGGATCCATGAGCAAAAAG
RACE Kit Fw 1	CCCCAAAACAAGCAGCCTTCCAACTA
RACE Kit Fw 2	TTCTCCTCAGACTCTGTCCCTGGTAA
M13 Fw	AGCATCAAATGGCGTGGAGA
M13 Rev	AGCATCAAATGGCGTGGAGA
Total pA Fw	CGGACAACTGGGCTCCTTCT
Total pA Rev	GCCTTGTCGTTGGAGGTGTT
pA2+3 Fw	CTCCCCAACCCCCTCATCTA
pA2+3 Rev	GGTTTACTAGGCGCCCCTTT
pA3 Fw	TCTGATCAAGGGAGAGGCCA
pA3 Rev	AGCATCAAATGGCGTGGAGA
18S Fw	GCAGAATCCACGCCAGTACAAGA
18S Rev	CCCTCTATGGGCCCGAATCTT
PTB Fw	GTTGGGTCGGTTCCTGCTAT
PTB Rev	CCGTCCATGGCACACAGAG
SRSF3 Fw	ACGGAATTGGAACGGGCTTT
SRSF3 Rev	CAAAGCCGGGTGGGTTTCTA
SRSF6 Fw	TTACGAGCTGAACGGCAAGG
SRSF6 Rev	CTTCCGTAGCTGTAGCCGTC
UV-XL Sequence 1 Fw	GTAATACGACTCACTATAGGGGCGGCGTCTCAGTGAAATC
UV-XL Sequence 1 Rev	GATCCGCAGGGGTGGATGCT
UV-XL Sequence 2 Fw	GTAATACGACTCACTATAGGGCTGACCCACAGCGTCACCCC
UV-XL Sequence 2 Rev	CTCTGACTCTTGGGTTAGAA
UV-XL Sequence 3 Fw	GTAATACGACTCACTATAGGGCAGCCAGTGTCTCCCATCA
UV-XL Sequence 3 Rev	TTTTAAGTAGGCCATGGTTGCT

Table 3. Name and sequence of all the primers used during the experimental work of this thesis. All primers from Sigma- Aldrich.

Results

CD5 3' end mapping

The first step of this work was to map the 3' ends of CD5 by 3' RACE and confirm previous results obtained with MOLT-4 cells [104]. Two pA sites were already identified in PolyA_DB (UCSC Genome Browser), being the most downstream one correspondent to the CD5 mRNA sequence NM_014207.3 at NCBI, used as reference for the primer design and alignments.

The Jurkat E6.1 cell line, a human T cell lymphoblastic cell line originated from an acute T cell leukemia, was used to perform RT-PCR using the poly(dT) adaptor primer (pink box in Fig. 4) which confines the amplification only to mRNA sequences with a pA tail. The first 3' RACE reaction was performed using a universal anchor primer, which hybridizes with the adaptor sequence of the poly(dT) adaptor primer, and a CD5 specific forward primer complementary to the most 3' region of the CD5 coding region (see *Material and Methods*, primer RACE Fw1). A nested PCR was performed

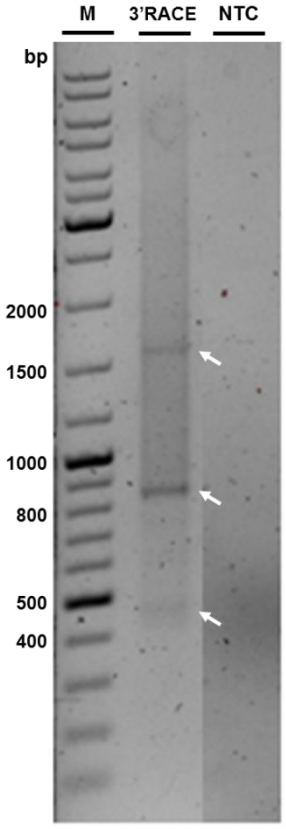


Figure 3. CD5 3' RACE in Jurkat cells. Representative gel electrophoresis of the CD5 nested 3' RACE using Jurkat cells mRNA cDNA; The white arrows indicate three CD5 pA isoforms: pA1 ~ 430 to 480 nt; pA2 ~ 850 nt; pA3 ~ 1530 to 1600 nt; M – DNA ladder; NTC – No template control.

using the same anchor primer and a CD5 specific forward primer on the beginning of the 3' UTR (see *Material and Methods*, primer RACE Fw2, and Fig. 4, blue box). The products were separated by gel electrophoresis (Fig. 3) and three bands are observed. The two longest ones have the expected sizes for the two pA sites reported in the PolyA_DB, and all three bands have the expected sizes of the mRNA isoforms identified previously in the Molt-4 cell line [104].

The PCR product of the nested 3' RACE reaction was cloned and sequenced, and the pA sites identified are represented in Fig. 4 (red arrows). All the five pA sites correspond to the usage of PAS already described in Molt-4 [104] (empty orange boxes). Also depicted in Fig. 4 are the more upstream PAS identified in Molt-4 [104] and the more downstream PAS associated with the longer isoform reported in PolyA_DB, both isoforms not obtained in this work (orange full boxes). The last isoform was also not sequenced in the Molt-4 work. The PAS associated with each pA site were identified by visual inspection of the sequence, according to spacing and sequence requirements (15-30 nt upstream of the pA site and hexamer AATAAA or variant) [42]. The different PAS were clustered in three major groups – pA1, pA2 and pA3.

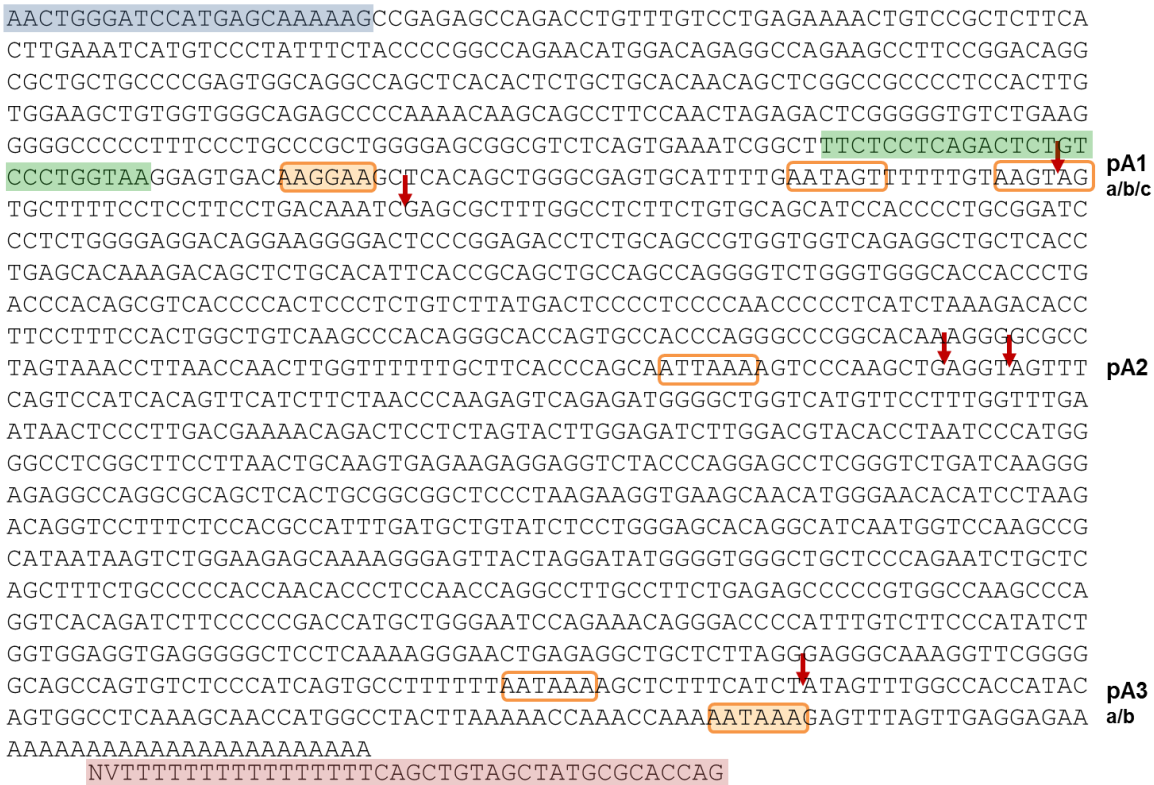


Figure 4. CD5 3' UTR. cDNA sequence correspondent to the CD5 3' UTR from NCBI (NM_014207.3), with the pA sites detected by 3' RACE/sequence in Jurkat cells represented by red arrows, and the correspondent PAS represented by empty orange boxes. Full orange boxes indicate the PAS associated with the pA sites reported in Molt-4 cells (pA1a) and in the PolyA_DB (pA3b). Blue box - RACE Fw2; Green box - RACE kit Fw2; Pink box - Poly(dT) adaptor primer.

The pA1 group includes the shortest isoform (pA1a) identified in Molt-4 but not in Jurkat cells, and pA1b and a, both sequenced in Jurkat. All three signals are non-canonical and predicted weak PAS, but, as referred by Domingues, 2010, this can explain the close proximity between them, as a safety mechanism to ensure cleavage and polyadenylation in this region [104]. The pA2 PAS, ATTAAA, which is the second most used [44], was associated with two different cleavage sites and also with the upstream PolyA_DB reported pA site, referred above. Finally, the pA3 group includes the pA3b PAS, associated with the downstream pA site deposited in the PolyA_DB database, and the pA3a PAS, both presenting the canonical sequence, AATAAA. For simplicity, the three groups of pA isoforms will be referred as pA1, pA2 and pA3.

The Jurkat cell line was the chosen biological model, because of the convenience of using a cell line comparative to primary cells. But in order to strengthen our results, 3' RACE was also performed in human primary T cells, in order to investigate if the three major groups of CD5 mRNA isoforms identified in Jurkat cells were also present in these cells. Also, it has been reported that proliferative cellular states such as induced pluripotent stem cells [54] and, particularly relevant for this work, activated T cells [53], were associated with general shortening of the 3' UTR length. Therefore, 3' RACE was also performed in PHA activated primary T cells in order to compare the CD5 mRNA isoform pattern in these cells comparing to resting T cells. These 3' RACE reactions were performed using the SMARTer^{TM} RACE cDNA Amplification kit (Clontech), with the RACE kit Fw1 (first reaction) and RACE kit Fw2 (nested reaction) primers (green box in Fig. 4). The nested reaction products were separated by gel electrophoresis and a representative picture is depicted in Fig. 5, where the same three major pA isoforms can be seen in all three cases, without significant differences between the two cell types and the different cellular conditions, resting and activated. The band correspondent to the pA1 isoform appears as a large diffuse band duo to the different isoforms composing it and the longer extension time applied to amplify the longer isoforms.

All the results from the 3' RACEs taken together indicate that the CD5 APA mRNA isoforms pattern is constant between the two different human leukemia cell lines, Jurkat and Molt-4, and human primary T cells, and between activated and resting primary T cells. Although is clear that the three major isoforms are expressed in all conditions, the 3' RACE is not a quantitative method, for which we cannot infer if the expression of the different isoforms is maintained or varies.

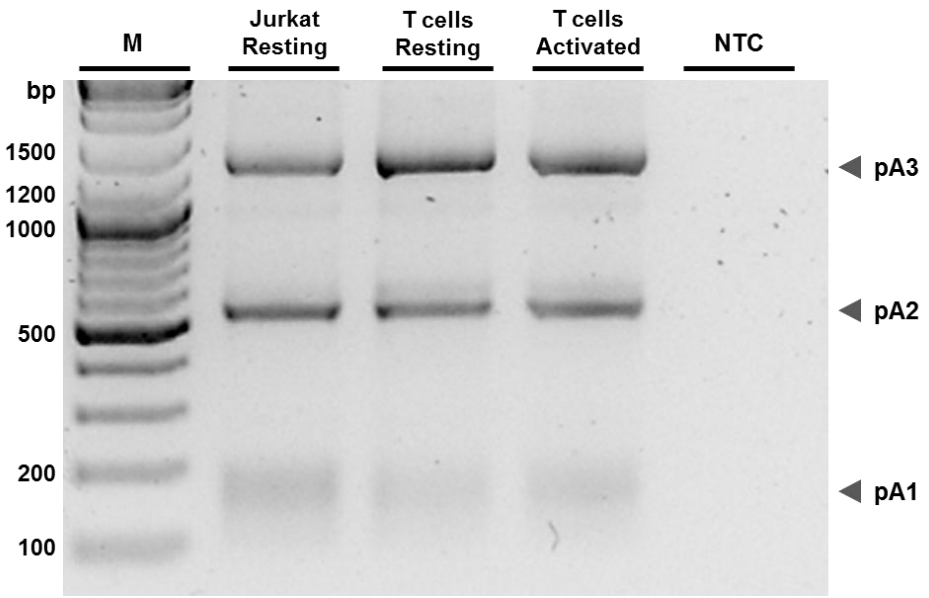


Figure 5. CD5 3' RACE in Jurkat and primary T cells. Representative gel electrophoresis of the CD5 nested 3' RACE using resting Jurkat cells, and resting and 24h PHA activated human primary T cells; The three major CD5 APA mRNA isoforms are present in all conditions and are indicated: pA1 ~ 110 to 160 nt; pA2 ~ 530 nt; pA3 ~ 1200 to 1280 nt; M - DNA ladder; NTC – No template control.

CD5 3' UTR conservation

The PhyloP method, from UCSC Genome Browser, was used to analyze the conservation pattern using 46 species of vertebrates (Fig. 6). Blue positive peaks represent conserved sequences while the red ones represent fast-evolving sequences. The CD5 3' UTR sequence is poorly conserved, yet presents some points of conservation, such as the regions of the three PAS which lack fast-evolving peaks, particularly the pA1 PAS that is located in a region with also strong peaks of high conservation.

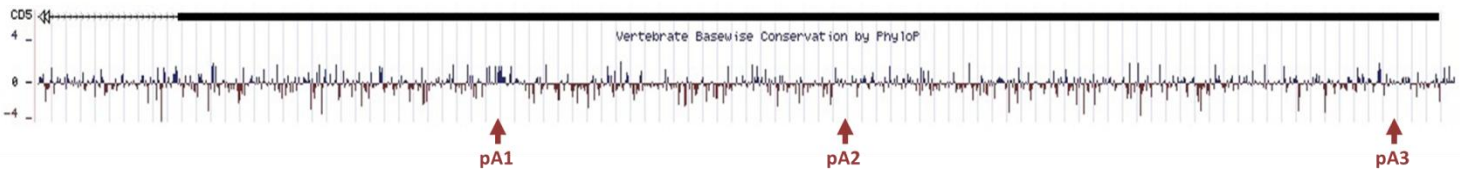


Figure 6. Human CD5 3' UTR conservation pattern of 46 species of vertebrates, from PhyloP in the UCSC Genome Browser. The positive blue peaks represent conserved regions and the negative red peaks represent fast-evolving regions. The three PAS (pA1, 2 and 3) are located in conserved regions, lacking fast-evolving peaks, and particularly the pA1 PAS presents strong positive conservation peaks.

In order to analyze with more detail the CD5 3' UTR conservation, the Geneious v4.8 software [105] was used to perform the alignment of seven CD5 genomic sequences from Ensembl of representative mammalian species: human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla gorilla*), pig (*Sus scrofa*), elephant (*Loxodonta africana*), dog (*Canis lupus familiaris*) and mouse (*Mus musculus*). The conservation percentages – pairwise percentage identity – are obtained from the scoring of each pair of bases of the same column, divided by the total pair number (Geneious 7.0 Manual at www.geneious.com).

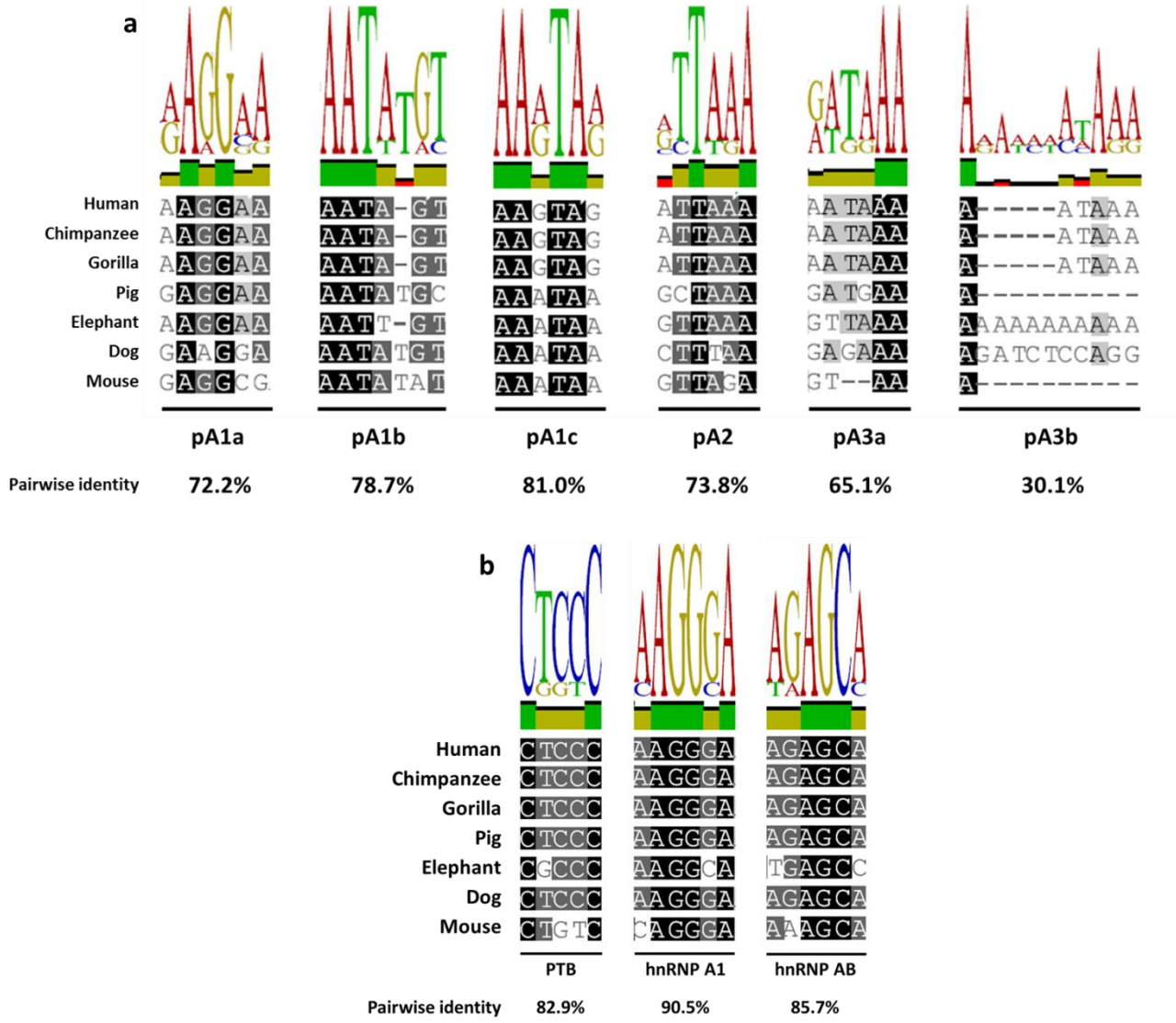


Figure 7. Human CD5 PAS and RBPs binding sites conservation in the CD5 3' UTR, with pairwise % identity indicated. **a**. The six PAS alignment using the Geneious v4.8 software with seven mammalian species CD5 genomic sequences. The pA1b and c PAS show the highest conservation and the pA3a and b PAS show the lowest conservation. **b**. Conservation of PTB and hnRNP A1 predicted binding sites, as examples of several high conserved sequences throughout the CD5 3' UTR.

The alignment of the six PAS in Fig. 7a clearly shows a higher conservation of the pA1 and pA2 PAS comparing to the pA3 PAS. The pA1b and c are the most conserved PAS, with 78.7% and 81.0% pairwise identities, respectively, while pA1a and pA2 present 72.2% and 73.8%, respectively. The least conserved PAS are pA3a and b, with only 65.1% and 30.1% identity. Moreover, the regions surrounding the pA1 PAS are strongly conserved, and the fact that these are predicted weak signals may be overcome by the strength of the up and downstream elements, USE and DSE.

Other specific sequences throughout the CD5 3' UTR show high conservation, including RBPs binding sites predicted by the SFmap online tool (described in *Predicted RBPs binding sites*), such as PTB or hnRNP A1 (Fig. 7b), and other regions that were not identified but may be important regulatory sequences.

CD5 APA isoforms expression

To further characterize the three major CD5 mRNA isoforms, their expression was measured by RT-qPCR in human primary T cells and in the E6.1 Jurkat cell line. Since the three isoforms do not differ in their sequences except for the extra length of the 3' UTR, the primers were designed targeting these areas. Therefore, the *Total pA* primer pair is located between exon 2 and 3 and amplifies all the CD5 mRNA isoforms, i.e., total mRNA levels, the *pA2+3* primer pair is located between the pA1 and pA2 PAS and amplifies the pA2 and pA3 isoforms, and the *pA3* primer pair is located between the pA2 and pA3 PAS and amplifies the longer isoform (Fig. 8).

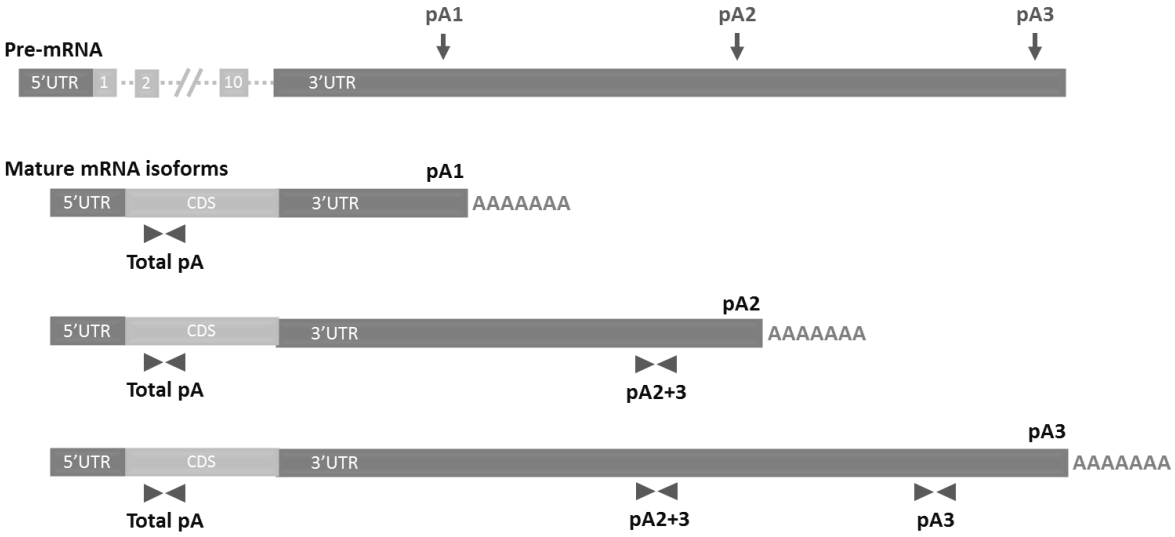


Figure 8. Schematic representation of the CD5 pre-mRNA and mature mRNA isoforms, with the designed primer pairs indicated. Dotted lines - introns; Dark boxes – 5' and 3' untranslated regions; Light boxes – exons and coding sequence (CDS); Vertical arrows - pA signals; Horizontal arrow heads - forward and reverse primers.

The relative expression (see *Material and Methods*) of the CD5 APA mRNA isoforms was measured in human primary T cells and Jurkat cells, in resting state. The results include three individual experiments for each cell type.

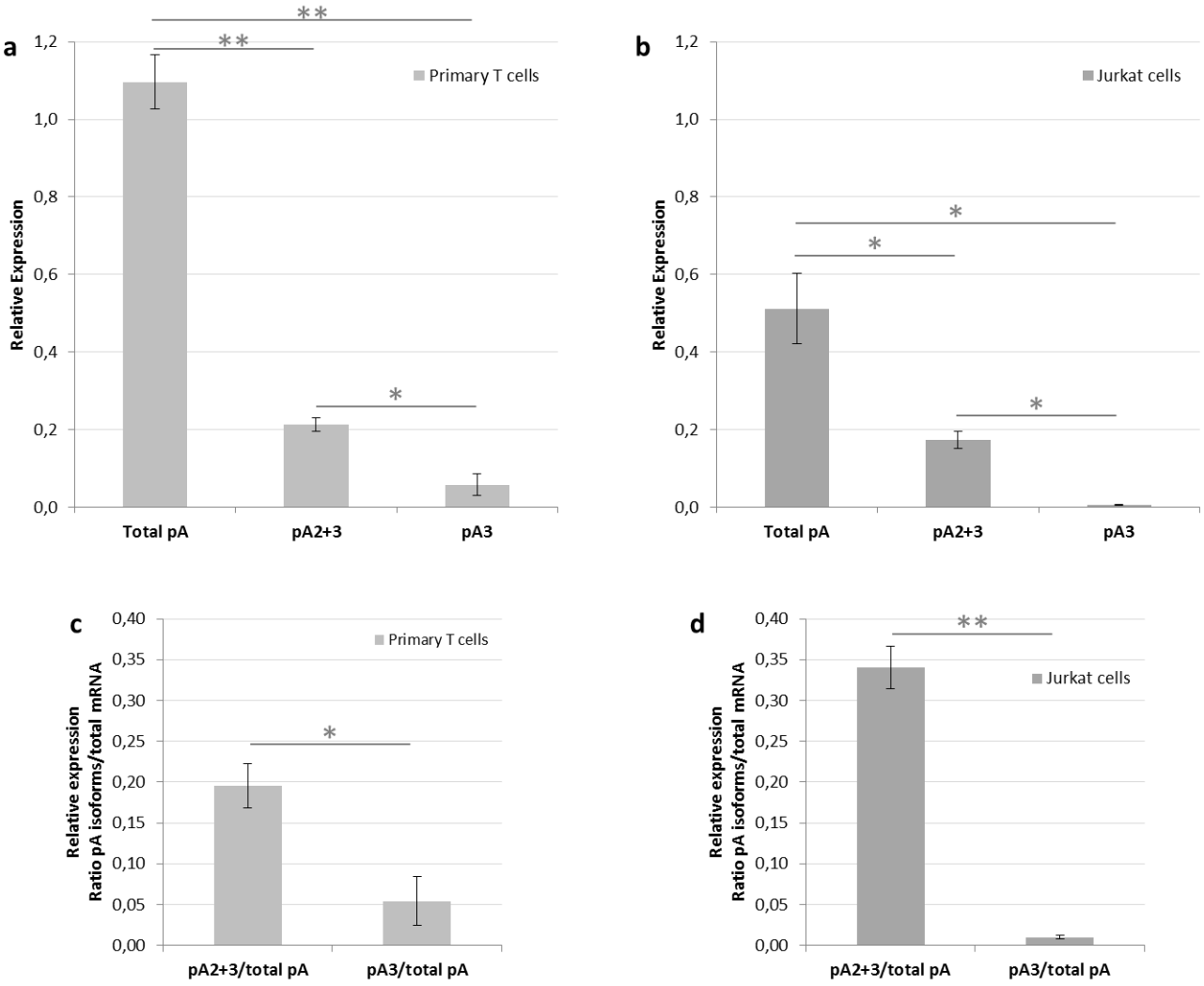


Figure 9. The RT-qPCR results show that the CD5 pA1 mRNA isoform is the most expressed and the pA3 is the least, both in primary and Jurkat T cells. **a**. Relative expression of the CD5 APA mRNA isoforms, in resting primary T cells. **b.** Relative expression of the CD5 APA mRNA isoforms, in resting Jurkat cells. **c**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in primary T cells. **d.** Ratio as described in **c.**, in Jurkat cells. * $p \leq 0.05$, ** $p \leq 0.01$.

Primary T cells showed a strong difference between the total mRNA expression and the two longer isoforms (pA2 and pA3 together represent ~20% of the total mRNA) for which it can be concluded that the pA1 isoform is the most expressed (Fig. 9a and c). Also, the expression of pA2 and pA3 isoforms together is four times higher than the pA3 individual, which indicates that the last isoform is the least expressed of the three representing slightly above 5% of the total mRNA. In Jurkat cells, the levels of total

mRNA and pA2+pA3 are not so disparate as in primary T cells, but the pA1 isoform is still the most highly expressed one, since pA2 and pA3 together represent less than 35% of the total mRNA (Fig. 9b and d). As for the pA3 isoform, it only represents ~1% of the total mRNA and less than 3% of pA2 and pA3 together.

Altogether, these results showed that the pA1 isoform is the most expressed of the CD5 APA mRNA isoforms, followed by pA2. The pA3 isoform presented very low levels, so the more significant difference between the two cell types is the proportion of pA1 comparing to pA2.

CD5 APA and T cell activation

As mentioned in *Introduction*, CD5 is an important regulator of the T cell activation. Therefore, one of the aims of this study was to investigate if this process was associated with the CD5 APA. First, the activation with PHA for 24h in Jurkat cells was confirmed by flow cytometry using an anti-CD69 antibody as activation control. This molecule was described as an early lymphocyte activation marker having its peak at 24h with a PHA activation [106]. The surface protein levels of CD5 were also measured in resting and activated cells, in order to compare these results with previous studies [104, 107].

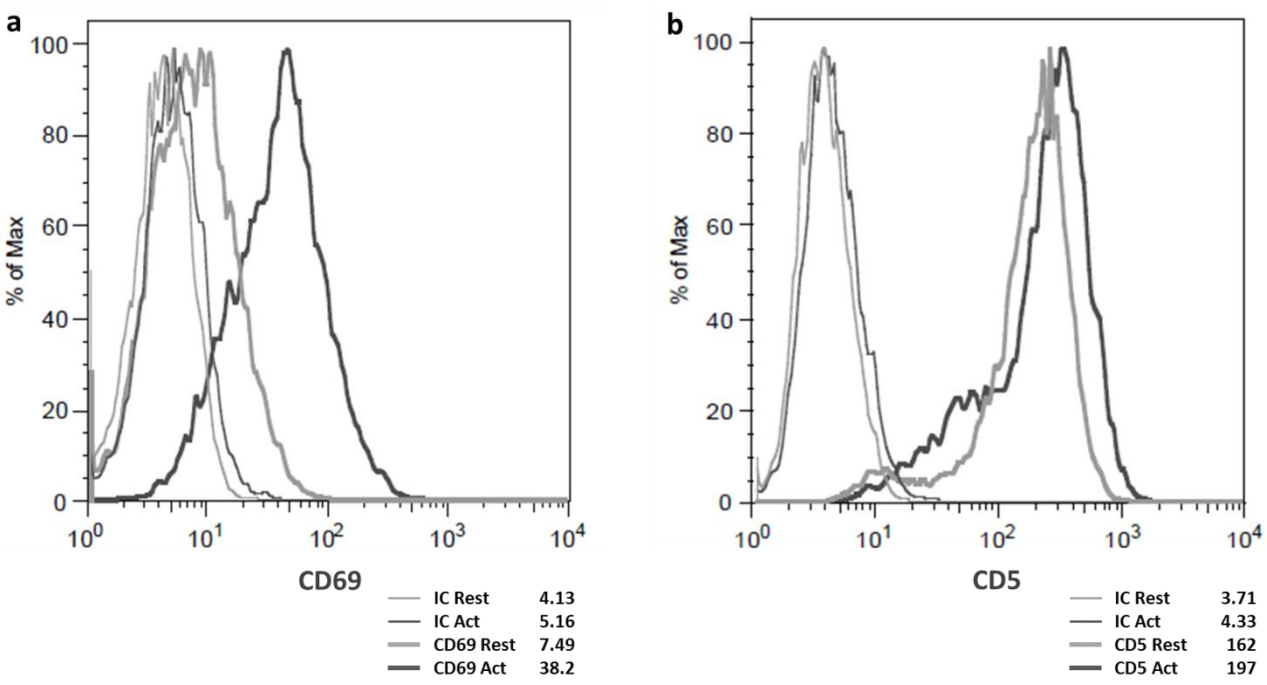


Figure 10. CD5 protein surface levels increase upon activation. Representative graphics of one of the three independent flow cytometry experiments performed in resting and 24h PHA activated Jurkat cells, with fluorescently labeled anti-CD69 (**a**) and anti-CD5 (**b**) Ab. Results analyzed using the geometrical mean values (tables below the graphs). IC refers to the isotype controls of the respective Ab. Both CD69 and CD5 surface levels increased upon activation.

The results were analyzed through the geometric mean values, which translate the average intensity of the fluorescent emission of the fluorescently labeled antibodies, using the FlowJo v. 8.8.7 software. Fig. 10 depicts one of the three independent flow cytometry experiments. The three results showed an average increase of 4-fold (± 0.75) of the CD69 protein levels with PHA incubation for 24h, indicating that the Jurkat cells were being activated. As for CD5, an average increase of 18% (± 14%) of the CD5 surface levels was obtained after 24h incubation with PHA (representative graph in Fig. 10).

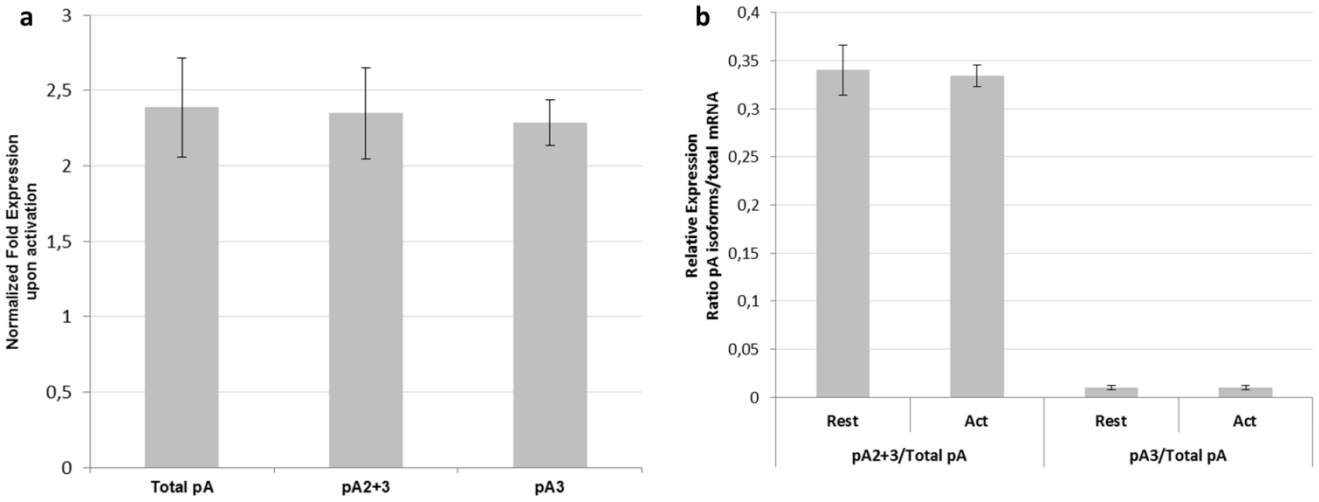


Figure 11. The RT-qPCR results show a general increase of the CD5 mRNA isoforms levels upon 24h PHA activation, in Jurkat cells. **a**. Normalized fold expression of the CD5 APA mRNA isoforms upon activation. All primers showed an increased expression of over 2-fold upon activation. **b**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in Jurkat resting and activated cells. The ratios did not differ upon activation.

After confirming the activation process, the CD5 APA mRNA isoforms levels were measured by RT-qPCR in Jurkat cells, in order to investigate the APA pattern in the two different cellular conditions – resting and 24h PHA activation. For these reactions the same three CD5 specific pair primers (see *CD5 APA isoforms expression*) were used, with 18S as the reference gene. Using the normalized fold expression (see *Material and Methods*) values, an increase of more than 2-fold of all the isoforms was observed upon activation (Fig. 11a). Analyzing the ratios of relative expression (see *Material and Methods*) of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, confirms that the increase is consistent between the different isoforms, indicating the same APA pattern in both conditions (Fig. 11b).

CD5 APA and chromatin state

It has been described that one of the levels of regulation of the APA mechanism is the epigenetic modifications, namely, by specific chromatin structures and histone modifications. To investigate if the CD5 APA could be influenced by alterations in the chromatin structure, human primary T cells were incubated with TSA (Trichostatin A) for 1h and the RNA from the nuclear fraction was extracted. Three independent experiments were performed with RT-qPCR reactions using the CD5 specific primer pairs described in *CD5 APA isoforms expression*. The TSA inhibits the class I and II mammalian histone deacetylases, promoting a global histone acetylation increase. Therefore, it "opens" the chromatin and allows it to be more easily transcribed. The normalized fold expressions (see *Material and Methods*) showed an increase of the pA2 and pA3 isoforms levels upon treatment with TSA, of about 20% for the pA3 isoform and 40% for the pA2 and pA3 together, while the total mRNA levels did not significantly changed (Fig. 12a). The pA2+3/Total pA and pA3/Total pA relative expression ratios also increased, supporting the results from Fig. 12a (Fig. 12b).

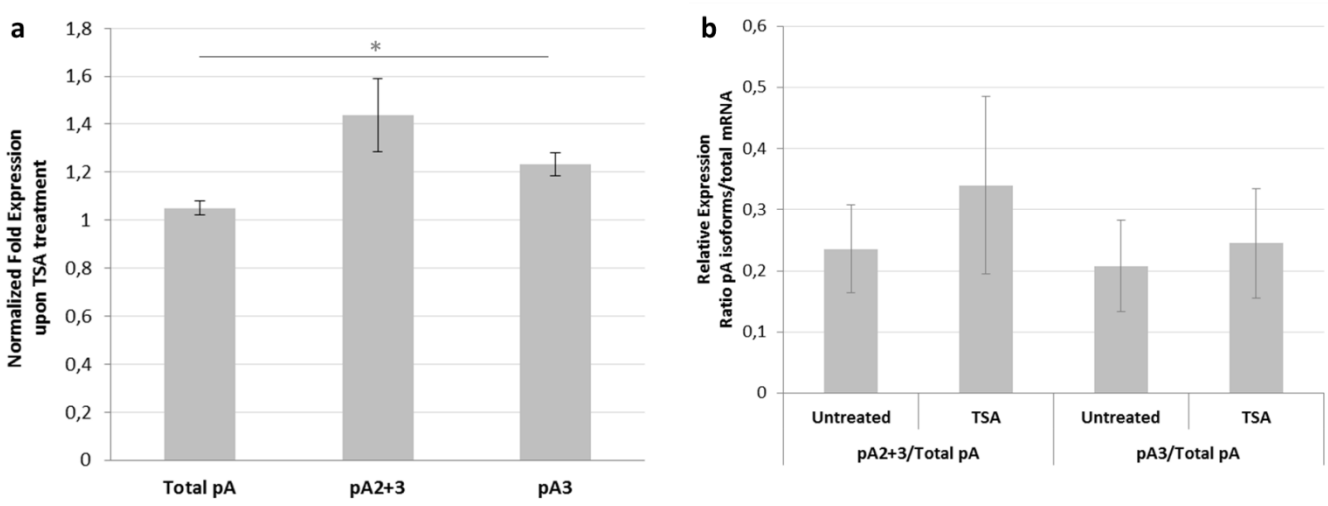


Figure 12. The RT-qPCR results show an increase of the pA2 and pA3 isoforms levels upon the TSA treatment, in human primary T cells. **a**. Normalized fold expression of the CD5 APA mRNA isoforms upon TSA treatment, relative to untreated cells. The *pA2+3* and *pA3* primer pairs showed an increased amplification of about 40 and 20%, respectively. **b**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in TSA treated and untreated primary T cells. Both ratios increased, displaying the same tendency of the results in Fig. 12a.

Predicted RBPs binding sites

Several RBPs have been associated with the 3' end processing and described as important factors to regulate the correct cleavage and polyadenylation [86]. Moreover, some have shown to influence the PAS alternative choice as its binding can promote or inhibit the usage of alternative PAS. Therefore, an important aim of this work was to identify RBPs capable of affecting the CD5 APA, in particular proteins with a role in the splicing mechanism. The first step was to search online databases for predicted RBPs binding sites in the CD5 3' UTR, for which we used the SFmap tool (sfmap.technion.ac.il) [108]. Its logarithm yields a score (0 to 1) based on the propensity of these binding sites to be clustered and on their high evolutionary conservation. There are several predicted binding sites for different RBPs throughout the CD5 3' UTR such as, SRSF1, 2, 3, 5 and 6, NOVA1, PTB, hnRNP A1, AB and H/F, Tra2α and β, or 9G8, suggesting that CD5 is a strong target of post-transcriptional regulation. To narrow down this list we focused on those for which we had knockdown resources, SRSF2, 3, 5 and 6, PTB, hnRNP A1 and AB, but only the PTB, SRSF3 and SRSF6 results are shown, since the other RBPs knockdown results were too preliminary at the time of this thesis submission.

CD5 APA mRNA isoforms expression and RBPs

The selected protein factors, SRSF2, 3, 5 and 6, PTB, hnRNP A1 and AB, were knockdown (KD) by synthetic siRNA transfection into Jurkat cells. siRNAs are 20-25 nt single stranded non-coding RNA molecules that target transcripts and inhibit gene expression by mRNA degradation or translation impairment [109]. Endogenously, long double-stranded RNA (dsRNA) molecules are cleaved by Dicer, which then directs the 20-25 nt dsRNA to be assembled into the RNA-Induced Silencing Complex (RISC). In this complex the two strands are separated and one of them is degraded while the other is the guide strand that is complementary to the target sequence [110]. Synthetic siRNA are designed as 20-25 nt dsRNA molecules overpassing the Dicer cleavage and being directly assembled on RISC.

Several important steps have to be taken into consideration when performing siRNA knockdown experiments. The most efficient transfection method should be selected for the specific target cells. For Jurkat cells, Amaxa® Nucleofection® is the transfection method of choice, already optimized in the lab. A careful siRNA design must be performed taking into account the possible unspecific targeting, targeting of all

isoforms and the strongest expression reduction. The last is often obtained using two to four siRNA for the same mRNA (General Design Guidelines by Life Technologies).

For the following experiments two siRNA molecules were designed for each RBP mRNA. As a negative control, a scramble sequence, which has no homology to any of the target mRNAs, was used. The incubation time and siRNA concentration were optimized for each target, and RT-qPCRs were performed to validate the knockdown in every independent experiment.

PTB

PTB binds to polypyrimidine tracts and several of its SFmap predicted binding sites show high scores (higher than 0.8 in a 0 to 1 range, from SFmap) and high conservation (pairwise % identities higher than 75% from the alignment of seven species using Geneious v4.8). Four individual experiments of siRNA KD of PTB in Jurkat cells were performed, with a ~70% average decrease of PTB expression (Fig. 13).

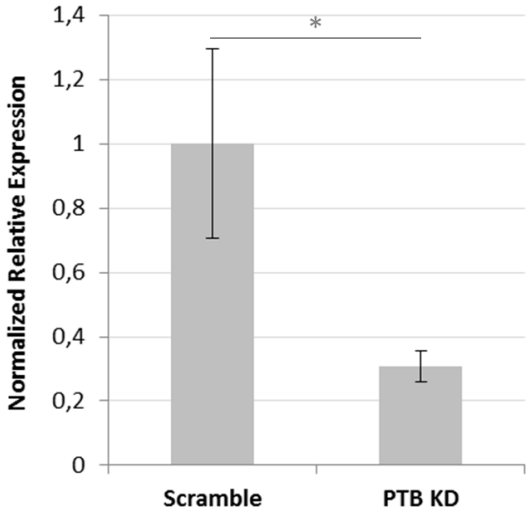


Figure 13. PTB expression decreased ~70% in the PTB siRNA KD, relative to Jurkat cells transfected with the scramble sequence. Relative expression of PTB of four individual experiments, using PTB specific primers and 18S as reference gene, normalized to the scramble value. * $p \leq 0.05$.

Fig. 14 shows the RT-qPCR results obtained from the PTB KD for the CD5 APA mRNA isoforms expression. The expression of the pA2 and pA3 isoforms together was increased ~14% upon PBT KD, while the total mRNA levels decreased ~16% (Fig. 14a). This decrease is mainly due to the decrease of the pA1 isoform, since the pA3 isoform levels do not showed significant differences. The wide standard deviation bar of pA3 is explained by the very low expression of this isoform that creates strong

variations in the normalized fold expression, even though the differences are small. The pA2+3/Total pA and pA3/Total pA relative expression ratios also increased (Fig. 14b). The first ratio is the result of the decrease in the total mRNA levels and the increase of the longer isoforms (pA2 and pA3). The second ratio do not suffered significant changes since it is only the result of the decreased of the total mRNA levels.

These results show that PTB KD causes an increase of the pA2 isoform expression and a decrease in pA1.

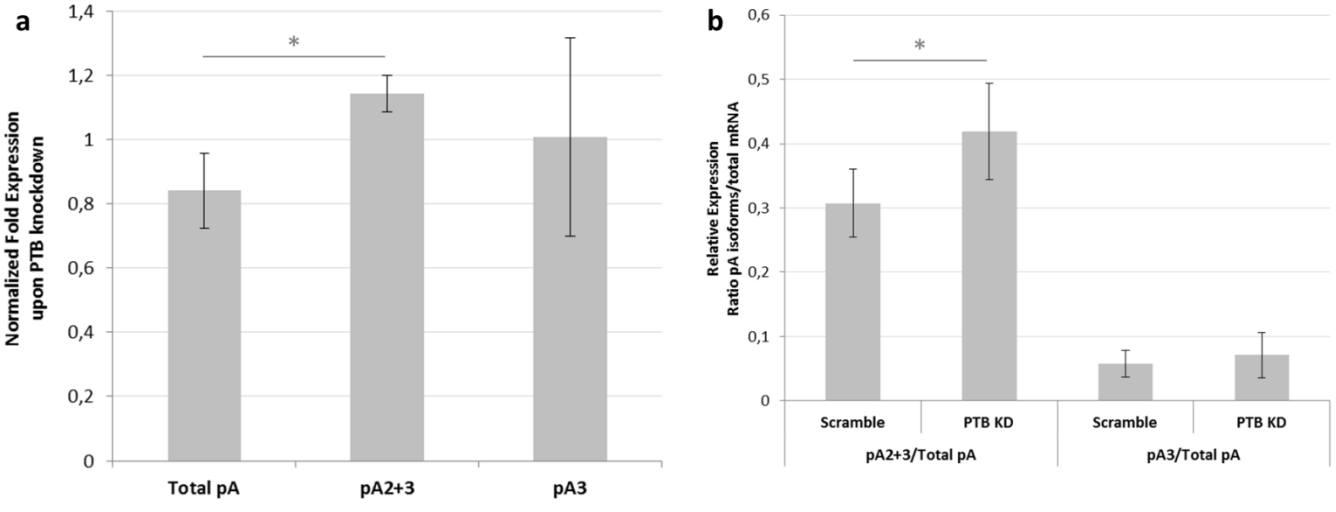


Figure 14. The RT-qPCR results show a decrease in the pA1 isoform expression and an increase in pA2 upon PTB KD, in Jurkat cells. **a**. Normalized fold expression of the CD5 APA mRNA isoforms upon PTB KD, relative to scramble transfected cells. The total CD5 mRNA level decreased while the expression of pA2 and pA3 isoforms together increased. **b**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in PTB KD and scramble transfected cells. Both ratios increased in PTB KD cells. * $p \leq 0.05$.

SRSF3

The SRSF3, or Srp20, binds to polypyrimidine tracts as PTB, but also to similar sequences occasionally containing As or Gs instead of Us. In the CD5 3' UTR the predicted binding sites for SRSF3 are uniformly distributed through its entire length, of which some present high scores (higher than 0.8 in a 0 to 1 range, from SFmap) and high conservation (pairwise % identities higher than 75% from the alignment of seven species using Geneious v4.8). Three individual experiments of siRNA KD of SRSF3 were performed, with a ~66% average decrease of SRSF3 expression (Fig.15).

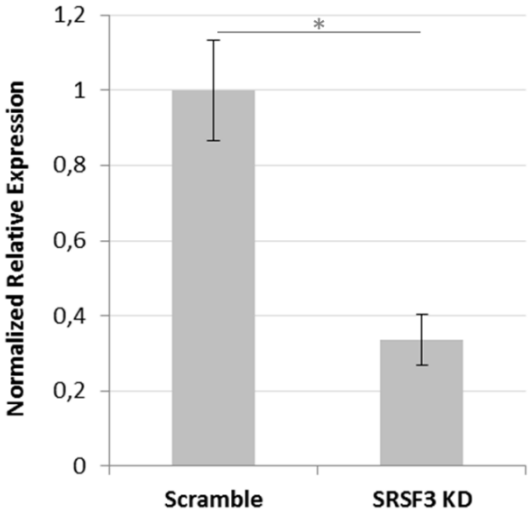


Figure 15. SRSF3 expression decreased ~66% in SRSF3 siRNA KD, relative to Jurkat cells transfected with the scramble sequence. Relative expression of SRSF3 of three individual experiments, using SRSF3 specific primers and 18S as reference gene, normalized to the scramble value. * $p \leq 0.05$.

The KD of SRSF3 resulted in an increase of about 70 to 90% of all primer pairs amplification levels (Fig. 16a). Also, the pA2+3/Total pA and pA3/Total pA relative expression ratios showed no significant differences in SRSF3 KD and scramble transfected Jurkat cells (Fig. 16b). These results indicate that SRSF3 inhibits CD5 expression in a general manner, without having a specific effect in the APA pattern.

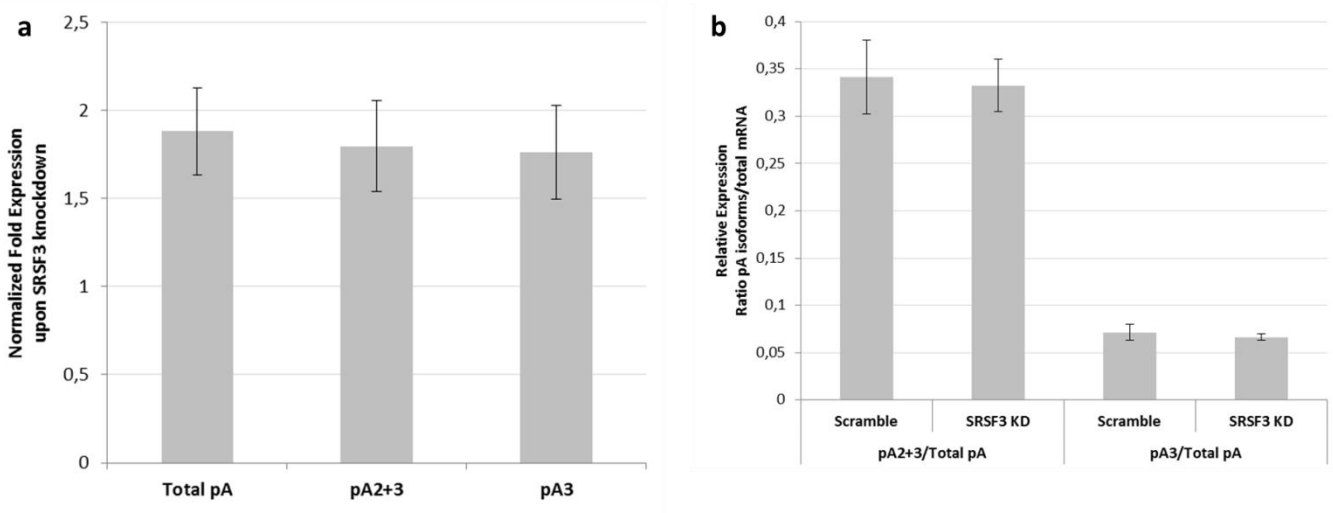


Figure 16. The RT-qPCR results show an increased in the CD5 total mRNA levels without a specific effect in the APA pattern upon SRSF3 KD, in Jurkat cells. **a**. Normalized fold expression of the CD5 APA mRNA isoforms upon SRSF3 KD, relative to scramble transfected cells. All primer pairs showed increased amplifications of about 70 to 90%. **b**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in SRSF3 KD and scramble transfected cells. Neither of the ratios presented significant differences.

SRSF6

The SRSF6, or Srp55, binding sites are specific hexamers, Y-R-C-R-K-M (see *Abbreviations*). The CD5 3' UTR have only a few SRSF6 binding sites and two of those are located near the pA1 PAS with a 0.86 score (in a 0 to 1 range, from SFmap). Six independent experiments of siRNA KD of SRSF6 were performed. However, the efficiency of KD varied from 20 to 95% and therefore, the data must be interpreted with caution. Four experiments were selected to analyze the CD5 mRNA APA isoforms, in which the average decrease of SRSF6 expression was ~62% (Fig. 17).

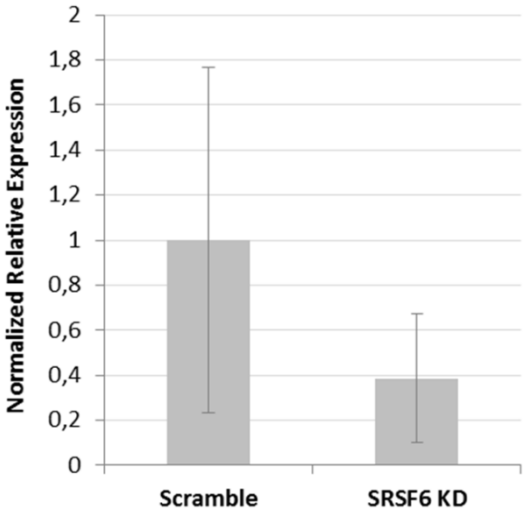


Figure 17. SRSF6 expression decreased ~62% in SRSF6 siRNA KD, Jurkat cells transfected with the scramble sequence. Relative expression of SRSF6 of three individual experiments, using SRSF6 specific primers and 18S as reference gene, normalized to the scramble value.

The expression level of the pA2 and pA3 isoforms together showed a ~25% increase, while the total CD5 mRNA decreased ~15% (Fig. 18a). These results taken together indicate that the pA1 isoform expression individually decreases. Whether the *pA2+3* increase is due to pA2 or/and pA3 increased levels, is a conclusion that cannot be confirmed by these data, as *pA3* presents a very wide standard deviation bar (Fig. 18a). Both pA2+3/Total pA and pA3/Total pA relative expression ratios increased in PTB KD cells, due to the decrease of the total mRNA levels and the increase of the pA2 and pA3 isoforms (Fig. 18b). An interesting relation between the SRSF6 KD percentage of all the six experiments and the pA3 normalized fold expression levels was identified: the more efficient the KD, the higher the pA3 increase (Fig. 18c). This effect is clearer in KD % above 60.

These results show that SRSF6 KD causes a decrease of the pA1 isoform expression and seems to lead to increase expression of the pA3 isoform.

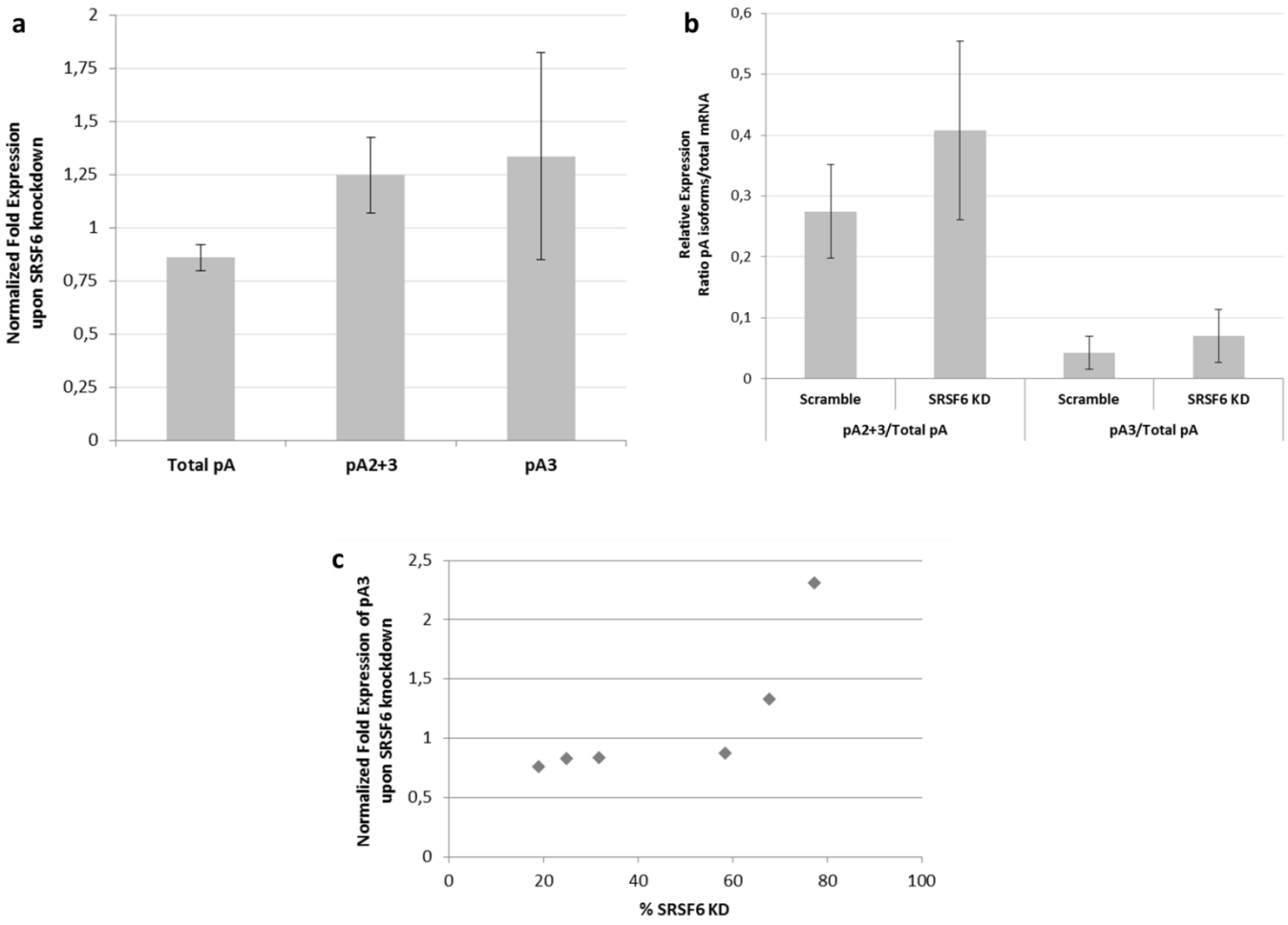


Figure 18. The RT-qPCR results show a decrease of the pA1 isoform and an increase of pA3 upon SRSF6 KD, in Jurkat cells. **a**. Normalized fold expression of the CD5 APA mRNA isoforms upon SRSF6 KD, relative to scramble transfected cells (n=4). The *pA2+3* levels increase while total CD5 mRNA levels decrease. **b**. Ratios between the relative expression values of the *pA2+3* and *pA3* primer pairs and the *Total pA* pair, in SRSF6 KD and scramble transfected cells (n=4). Both ratios tend to increase upon SRSF6 KD. **c**. Relation between the normalized fold expression of pA3 and the SRSF6 KD efficiency % (n=6). The higher the KD %, the stronger the increase of the pA3 isoform levels.

Specific RBPs bind the CD5 3' UTR

We have identified correlations between the knockdown of PTB and SRSF6 and differences in the CD5 APA pattern. To investigate if these results correspond to the binding of these specific factors to the CD5 3' UTR, UV cross-linking assays were performed. This method consists in the cross-linking between proteins and nucleic acids through formation of covalent bonds by ultraviolet irradiation (UV). Three Jurkat cDNA sequences around the CD5 3' UTR PAS (Fig. 19) were used as template to *in vitro* transcription with α-^{32}P radiolabeled nucleotides, Us and Cs. The radiolabeled RNA was incubated with nuclear extracts prepared from Jurkat cells, and UV

cross-linked, followed by RNase digestion. Running these reactions – inputs – in a SDS-PAGE gel will show all proteins that bind to the specific sequences used (see *Material and Methods*).

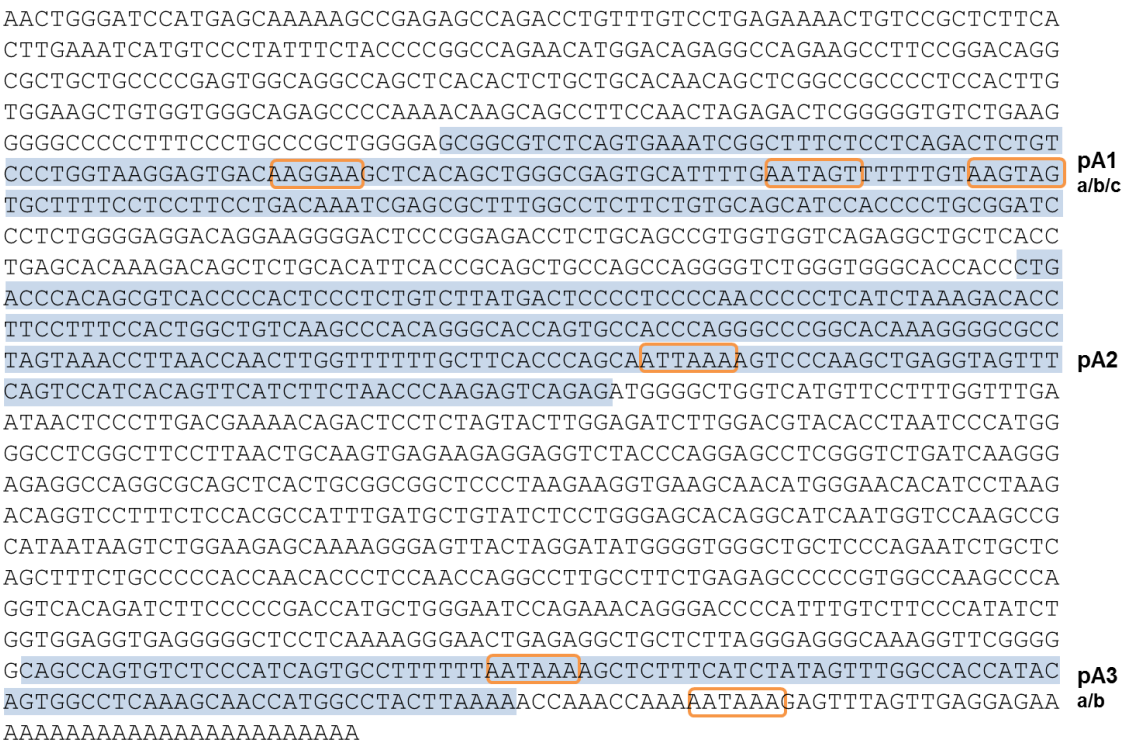


Figure 19. CD5 3' UTR with PAS indicated with orange empty boxes and sequences used for the UV cross-linking assays indicated with blue full boxes.

The sequences to perform UV cross-linking were selected based on the proximity to the PAS, at least 50 to 70 nt up and downstream of the PAS, and the presence of important conserved and high score binding sites beyond these limits (as the upstream sequence of the pA1 PAS – sequence 2). Fig. 20 represents the predicted RBPs binding sites for PTB, SRSF3 and SRSF6, and also for hnRNP A1 and AB, and SRSF1, 2 and 5, contained in the three transcribed sequences, 1, 2 and 3.

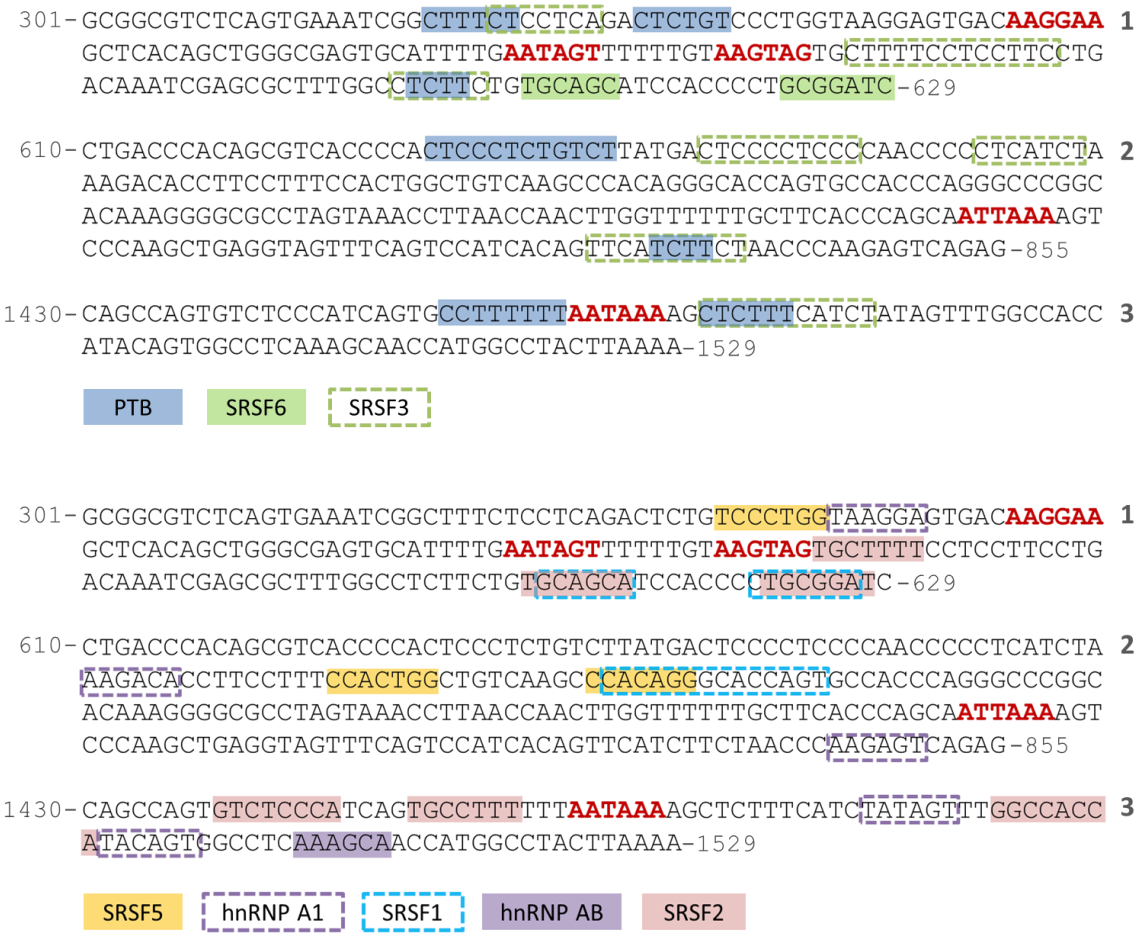


Figure 20. CD5 3' UTR cDNA sequences used as template for the radiolabeled RNA transcription (without the T7 promoter sequence represented) in the UV cross-linked assays. First and last nucleotides of the three sequences, 1, 2 and 3, are indicated, counting from the first nucleotide of the CD5 3' UTR. SFmap predicted binding sites for PTB, SRSF1, 2, 3, 5 and 6, and hnRNP A1 and AB, contained in the three selected sequences, are represented. Bold red characters represent the CD5 PAS.

From the input of the three different samples (representative gel in Fig. 21a) it was possible to detect some differences between them in the band pattern correspondent to the RNA-protein complexes. The band slightly above 37 kDa (E) can be identified in the sequences 1 and 3, but seems absent from 2; the band at 25 kDa (B) in sequence 1 does not appear in 2 and 3; the band at ~34 kDa (C) in sequences 1 and 2 is not present in sequence 3; the band of ~35 kDa (D), in sequence 3, is not present in sequence 1 and 2. These results suggest that different RBPs bind to different CD5 RNA sequences.

Part of the inputs were used to perform immunoprecipitation with an polyclonal anti-PTB antibody [111] (see *Material and Methods*). In these experiments, PTB was detected in sequence 1, but the same band was barely visible in sequence 2 and 3

(arrowheads in Fig. 21b). A strong background was obtained in the IP of sequence 1, and a very weak signal in the IP of sequences 2 and 3. This experiment must be repeated, although it can be affirmed that the predicted band corresponding to the RNA-PTB complexes is present in the three reactions.

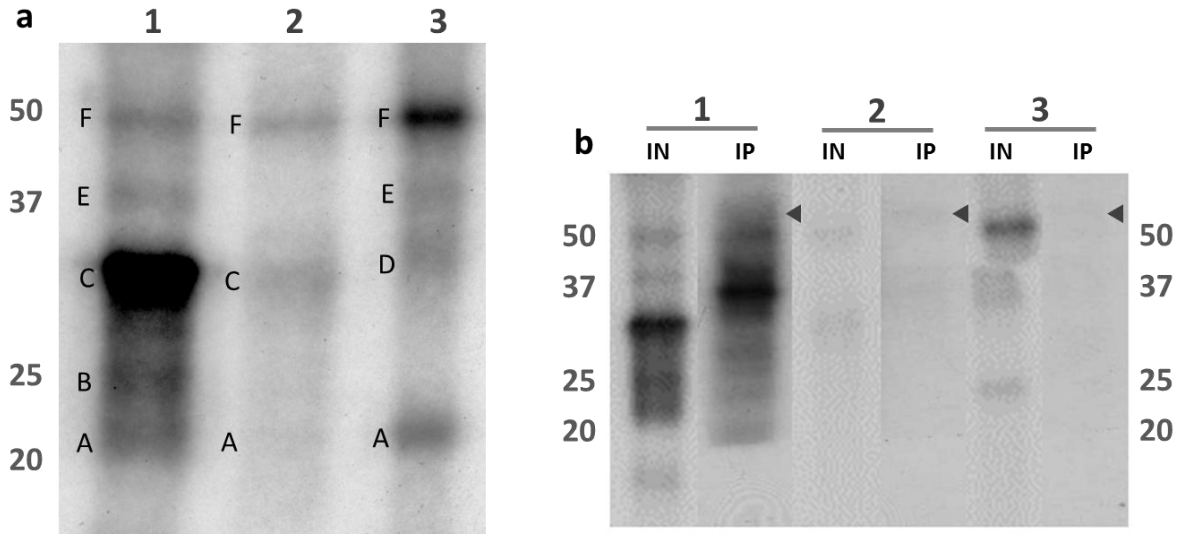


Figure 21. Several different RBPs, and in particular PTB, bind the human CD5 3' UTR, in Jurkat cells. **a**. Specific RBPs bind differently to the three PAS regions (Fig. 19 and 20). SDS-PAGE gel of the protein-RNA complexes of sequence 1, 2 and 3, in the UV cross-linking assays. Bands marked from A to F represent distinct proteins and are referred in the text. **b.** SDS-PAGE gel of the sequence 1, 2 and 3 inputs (IN) and correspondent PTB IP reactions (IP). The PTB IP showed faint bands above the 50 kDa mark that may be specific for PTB.

Discussion/Conclusion

The transmembrane glycoprotein CD5 functions as a regulatory molecule through its inhibitory effect in the TCR and BCR activity [21, 22, 31]. Its deregulation has been studied in the context of autoimmune processes and tumor immune response, and its role in human disorders raises the importance of studying this molecule, particularly at the expression regulation level. The 3' end processing is essential to yield a correct mature mRNA isoform, and the selection of different PAS, APA, plays a major role in the regulation of gene expression [51]. Alterations on the APA pattern, favoring the shorter or longer isoforms, are described in different cell types and cellular states [52-54]. APA has also been increasingly associated with other mRNA processing mechanisms (reviewed in *Introduction*) and an important aim of this work was to unveil some of these interactions. CD5 is a pan-marker of T cells, so the biological models chosen were human primary T cells and the E6.1 Jurkat cell line.

CD5 mRNA isoforms

The three major CD5 mRNA isoform groups, pA1, pA2 and pA3, already described in Molt-4 cells, were also shown to be present in Jurkat, and in resting and activated primary T cells. The specific isoforms in each group were identified by sequencing the 3' RACE products of Molt-4 [104] and Jurkat cells. The pA1 group includes three different isoforms, all using non-canonical PAS, of which the first one was only identified in Molt-4 cells. As for pA2 isoform, two distinct pA cleavage sites were identified in Jurkat cells different from the oneipredicted by PolyA_DB. However, all of them are defined by the same pA2 PAS which is the second most frequently used hexamer (AUUAAA) [44]. Finally, the pA3 group is composed by two different isoforms, each one using a different canonical PAS. Only the first PAS, pA3a, was identified in Molt-4 [104] and Jurkat cells, while the last one, pA3b, is reported in PolyA_DB.

Quantification of the CD5 mRNA isoforms by RT-qPCR, using the primer pairs described in *Results*, does not provide the individual expression of the three mRNA isoforms, except for the longest one, pA3. Nevertheless, it gives insight about the proportions between them, and showed that the pA1 isoform is the most expressed while pA3 is the least, both in Jurkat and human primary T cells.

The presence of the three pA1 putative weak PAS close to each other suggests that they may function together to induce transcription termination in the pA1 region

[104]. Furthermore, the analysis of the surrounding sequences of these PAS showed a strong conservation, suggesting that the auxiliary elements, USE and DSE, can contribute for the strength of this PAS group, a hypothesis indicated in [49]. By contrast, the pA3 region is the least conserved and, although presents the predicted strongest PAS hexamers (AAUAAA), it is the least expressed isoform. The increasing presumable strength of the PAS along the CD5 3' UTR (proximal to distal) is consistent with previous data that showed that non-canonical PAS are frequently proximal to the coding region while the 3' end pA sites usually use canonical PAS [44]. Furthermore, although pA1 is the most expressed, two other isoforms, with different target sequences for modulating factors, could account for additional posttranscriptional regulation, for example, in a tissue-specific manner.

CD5 APA in transformed cells

As mentioned before, it is widely accepted that the expression of APA isoforms change in different cellular conditions. For example, dividing cells have a tendency to use the shorter isoforms while differentiating cells use more frequently the longer isoforms [53, 54]. Oncogenic transformed cells are also associated with APA switch, favoring the usage of the shorter isoforms, proposed to be an miRNA evasion mechanism allowing increased protein levels [59].

Although pA1 and pA3 are the most and least expressed isoforms, respectively, in both Jurkat and human primary T cells, the proportions of CD5 APA isoforms between them are different. Primary T cells have a higher fraction of the pA3 isoform (~5%), but the two longer isoforms together, pA2 and pA3, represented only about 20% of the total mRNA levels. Jurkat cells, on the other hand, have a higher contribution of the pA2 isoform, as the pA2 and pA3 isoforms represent about 35% of the total mRNA with an insignificant contribution of pA3 (~1%).

The lower percentage of pA3 in Jurkat cells is in agreement with the expected lower levels of the longer isoforms in transformed cells, described by Mayr and Bartel [59]. Moreover, several putative miRNA target sites between the pA2 and pA3 PAS, such as mir-7, 124, 204 and 211, can be identified in online databases (microRNA.org, and TargetScan.org). mir-7 is reported to be expressed in T cells (unpublished data and microRNA.org) so it may regulate pA3 levels, although it may not be a relevant biological mechanism of CD5 APA regulation, due to its scarce expression. As for the proportion between the pA1 and pA2 isoforms, Jurkat cells tend to privilege the usage of the second PAS. Previous work demonstrated higher translation efficiency for the

pA2 isoform over pA1 and pA3 [104], which may justify the increased percentage of pA2 isoform in Jurkat cells in an attempt to optimize protein production.

CD5 APA upon T cell activation

An important cellular condition to be addressed when working with T cells is the activated state. To access if this condition can modulate the CD5 APA, cells were incubated with PHA, which triggers the immune response rendering them activated. In order to confirm the activation in Jurkat cells, the surface protein levels of CD69, an early T cell activation marker, were measured by flow cytometry. As expected, cells incubated with PHA for 24h showed an average 4-fold increase. In these experiments, the CD5 surface levels were also determined, showing a 20% increase in activated cells. A similar increase was reported in a previous work, after 24h incubation with CD3/CD28 stimulation, showing also a 2-fold increase after 48h, in Jurkat cells [104]. A second study using whole-blood samples showed no significant differences in the CD5 protein levels after 24h PHA activation [107].

Sandberg *et al*. reported an increased usage of the upstream PAS, and consequently a shortening of the 3' UTRs, in murine CD4+ T lymphocytes upon activation by CD3/CD28 stimulation [53]. It was suggested that this could be a widespread mechanism, raising the possibility that it could happen in CD5. Therefore, CD5 mRNA levels and APA pattern were accessed by RT-qPCR in resting and 24h PHA activated Jurkat cells. The total mRNA levels present a ~2-fold increase in activated cells, which was expected [1] and in line with the higher CD5 surface protein levels showed by flow cytometry. As for the CD5 APA pattern, the usage of the different isoforms was not altered. There are at least three facts that can explain these contradictory results: the experiments were performed in malignant transformed cells, Jurkat cells, in which the molecular response upon activation can be different from primary T cells; the T cell activation process is different when triggered by different methods (PHA or CD3/CD28 stimulation); and, for CD5 in particular, it can be a specie-specific process of mouse cells but not human.

CD5 APA and chromatin

Although APA is an RNA processing mechanism, its modulation and regulation starts at the DNA level. Alteration of the chromatin structure modulated by histone modifications is an example of an APA regulation level (as described in *Introduction*).

Different chromatin condensation states can facilitate or prevent access of the transcription machinery, having an impact on transcription kinetics [112]. Also, an association between APA and RNAP II kinetics has been demonstrated, as a "slower" reading speed favors proximal PAS usage [80]. In order to investigate if the CD5 APA pattern could be influenced by histone modifications, Jurkat cells were incubated with TSA, which promotes a general "opening" of chromatin by increasing histone acetylation. The results revealed an increase in the usage of the pA2 and pA3 isoforms while maintaining the total mRNA levels, which indicates that the longer isoforms are favored over the pA1 isoform. These data suggest that pA2 and pA3 signals are contained in a chromatin region with generally lower levels of acetylation than the pA1 signal, and that the increase of pA2 and pA3 mRNA levels by TSA treatment would allow them to be more easily accessed and recognized.

Two complete CD5 3' UTR maps of the acetylation mark H3K9ac (acetylation of histone H3 Lysine 9), a transcription activation mark [113], were accessed through the UCSC Genome Browser, originated by ChIP-seq (Chromatin Immunoprecipitation followed by high-throughput DNA sequencing) data from ENCODE/Stanford/Yale/USC/ Harvard (Fig. 22). These maps are from the K562 (human chronic myelogenous leukemia) and NT2-D1 (human embryonal pluripotent carcinoma) cell lines, since no information on primary or T cell lines could be obtained. Nevertheless, the two cell lines seem to reveal the same tendency: pA2 and pA3 PAS are located in less enriched H3K9 acetylation regions, comparing to pA1 region.

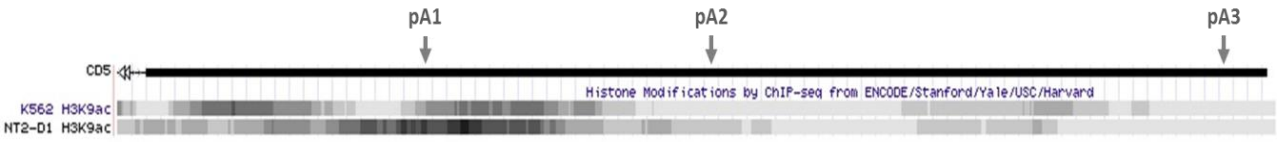


Figure 22. Maps of the H3K9ac modification in the CD5 3' UTR in K562 and NT2-D1 cells, with the CD5 PAS indicated. The grey intensity represents more (darker) or less (lighter) enriched sequences for the specific histone modification. Maps of ChIP-Seq data from ENCODE/Stanford/Yale/USC/Harvard (UCSC Genome Browser). The pA1 PAS is located in a region with higher levels of H3K9 acetylation than pA2 and pA3.

These acetylation maps indicate that, in a normal context, the pA1 region is more "open" and prone to be recognized by the transcription/3' processing machinery, which agrees with the fact that pA1 is the highest expressed CD5 isoform. Moreover, this hypothesis is supported by the TSA treatment experiments that showed an increase of the pA2 and pA3 levels (Fig. 23). Although all the results seem to correlate and suggest that histone acetylation can influence the CD5 APA, it will be important to characterize the general acetylation levels in T cells.

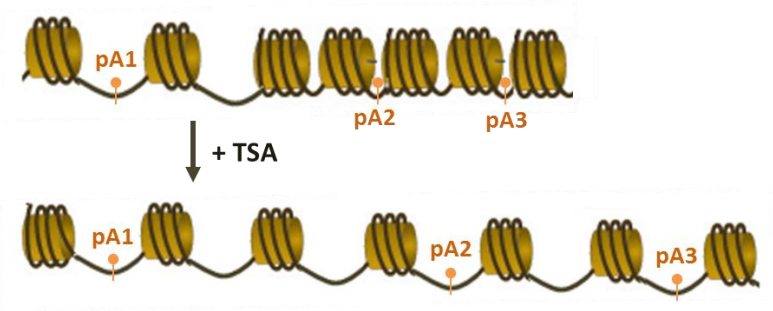


Figure 23. Schematic representation of the proposed model of the chromatin structure influence in the CD5 APA. The pA1 PAS region presents higher levels of acetylation than pA2 and pA3, being more frequently recognized and used. After TSA treatment, the chromatin "opens" the more condensate pA2 and pA3 PAS regions, by the inhibition of histone deacetylases, and their expression levels increase.

CD5 APA and RBPs

An increasing number of studies have associated alternative polyadenylation with protein factors commonly known for their functions in other processes (reviewed in *Introduction*). An important aim of this work was to identify RBPs that influence the CD5 APA, in particular proteins with splicing functions. As far as we know, CD5 only produces APA mRNA isoforms, not undergoing alternative splicing. This makes it a good candidate to study the role of splicing factors in APA, without co-occurrence of alternative splicing. The CD5 3' UTR contains a myriad of predicted binding sites for different splicing factors, such as SRSF1, 2, 3, 5 and 6, NOVA1, PTB, hnRNP A1, AB and H/F, Tra2α and β, and 9G8 (SFmap). Knockdown of SRSF6 and PTB indicated an influence of these protein factors in CD5 APA.

PTB

The CD5 3' UTR harbors several conserved predicted binding sites for PTB, in particular located in the surrounding regions of all three PAS. This protein has been associated with correct pre-mRNA 3' end processing of several genes, particularly by promoting the recruitment of polyadenylation factors (reviewed in *Introduction*). In the C2 gene it was shown that higher levels of PTB inhibit polyadenylation by competition with polyadenylation factors and therefore a narrow concentration interval is required for PTB's optimal effect. All reported data associates these effects to its binding to USE and DSE sequences.

Recently, a role in APA shift was also described for PTB [95]. In CD5, the knockdown of this protein resulted in an increased expression of the pA2 isoform over a decrease of pA1, with no significant changes for pA3. Therefore, it seems to act as a

stimulatory factor for the use of the pA1 PAS while having an inhibitory function on pA2 (Fig. 24). As already described, the pA1 PAS region, including USE and DSE, is the most conserved, suggesting a relevant function of these sequences. On one hand, it is possible that PTB more easily promotes the recruitment of polyadenylation factors to the pA1 PAS region, and when at low levels, the pA2 PAS is favored by its supposed higher processing efficiency (AUUAAA). On the other hand, PTB may be competing with polyadenylation factors for the binding to the USE or DSE of pA2 PAS, and its knockdown allows this isoform to be more expressed over pA1.

To confirm if PTB binds the CD5 3' UTR, a UV cross-linking assay was performed for the three regions around each PAS, followed by PTB immunoprecipitation. However, the experiment was not conclusive since the bands were too faint, for which further experiments will be required. Nevertheless, in the three sequences, a ~57 kDa band was visible, correspondent to the PTB molecular weight.

SRSF3

SRSF3 (Srp20), as PTB, has several conserved predicted binding sites in the CD5 3' UTR. Previously published results show that SRSF3 has a role in the 3' end processing, but different studies associate it with different mechanisms, either transcription termination, cleavage or polyadenylation. Its knockdown promoted an upregulation of CD5 mRNA levels of almost 2-fold, without a particular tendency for a specific isoform, suggesting an inhibitory role over CD5 expression. Since all previous studies reported a positive effect of SRSF3 in the 3' end processing, it would be interesting to study the effect of this factor in other genes and the mechanism behind this observation.

In the UV cross-linking assays a ~23 kDa band was possible to identify in all three sequences, correspondent to the SRSF3 molecular weight. In sequence 2 the band was barely visible due to the global weaker signal of this sample. Immunoprecipitation with a specific anti-SRSF3 Ab would confirm if this band is in fact SRSF3.

SRSF6

SRSF6 (Srp55) has few predicted binding sites in the CD5 3' UTR, specifically located near the pA1 PAS and between the pA2 and pA3 PAS. No previous studies have shown an involvement of Srp55 in the 3' end processing, at the cleavage or polyadenylation level. Its knockdown showed inconsistent results, both in the knockdown efficiency and the CD5 isoform quantification. Even though, it was possible to identify a tendency of pA1 decrease and pA2 and pA3 increase. Moreover, the

higher the percentage of knockdown, the higher the pA3 increase. Therefore, it may exert a stimulatory effect in the pA1 PAS while having an inhibitory function on pA3 (Fig. 24). The influence of this protein in APA is a new observation and should be further investigated in the future.

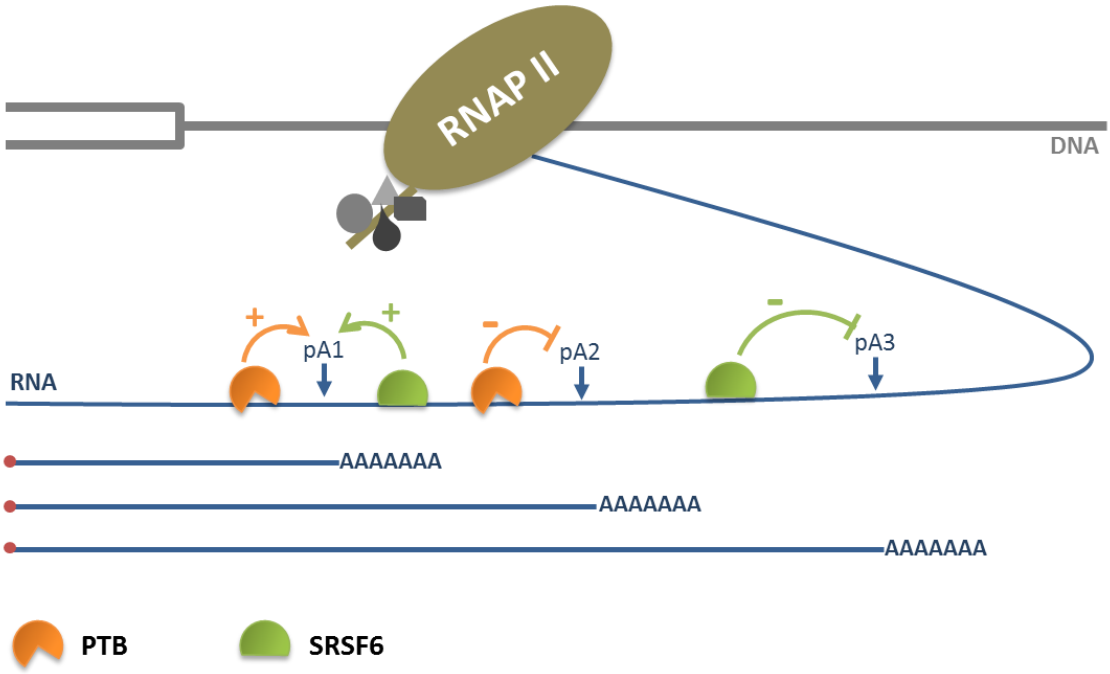


Figure 24. Schematic representation of the proposed effect of PTB and SRSF6 in the usage of the three CD5 PAS. PTB has predicted binding sites in the pA1 and pA2 proximity and plays an inhibitory regulation of pA2 and a stimulatory regulation of pA1. SRSF6 has predicted binding sites in the pA1 and pA3 proximity and plays an inhibitory regulation of pA3 and, as PTB, a stimulatory regulation of pA1.

Other RBPs

The UV cross-linking results provided an important support for the hypothesis that CD5 APA regulates CD5 expression. Not only showed that several RBPs bind to its 3' UTR, but also that they do it differentially, suggesting specific functions of these proteins in specific PAS regions. It is also important to bear in mind that these assays only accessed short sequences around the three PAS and that other conserved binding sites are located throughout the 3' UTR for all the proteins analyzed. Therefore, their function in CD5 APA may also depend on other regions of the 3' UTR or even other interactions between different proteins.

Considering the predicted binding sites from SFmap, their scores and conservation, accessed by Geneious v4.8, the possible identity of some of the proteins

detected in the UV cross-linking assays is proposed. The ~33 kDa band C, is present in sequence 1 and 2 but not 3, and in fact there are no predicted binding sites for SRSF1 (SF2/ASF), according to SFmap, in the last sequence 3. Band D, which is located slightly above band C (~35 kDa), seems to bind to sequence 3. This band could correspond to hnRNP A1, with ~35 kDa, which presents putative binding sites in the three sequences, although the sites in sequence 1 and 2 contain only one labeled nucleotide each (see Fig. 20 in *Results*), that may render visualization difficult. Searching for ~50 kDa proteins that may bind the CD5 3' UTR, NOVA1 stood out with several conserved and high score binding sites in the three sequences.

Taken together the results of this work, we were able to confirm that distinct cellular conditions, chromatin states and protein factors are able to induce differences in the proportions of the CD5 APA mRNA isoforms. This suggests that CD5 is regulated by APA, which in turn is modulated by several mechanisms that they may work together in a PAS specific manner.

Future work

It was shown that although the CD5 APA pattern is consistent between Jurkat and human primary T cells, the proportions of the different mRNA isoforms vary. The decrease in pA3 percentage may be associated with miRNA targeting, hence levels of mir-7 in the two cells will be accessed, as well as new miRNAs with CD5 3' UTR target sites and expression in T lymphocytes will be identified. Also, the mechanism explaining the higher proportions of pA2 isoform in Jurkat cells will be investigated, since it may unveil an important feature that could be tumor-specific.

As APA is generally known to be tissue specific, t would be interesting to identify the CD5 APA pattern in other cell types, in particular B1a lymphocytes, in order to understand if the pattern presented is T cell specific.

The CD5 APA isoforms expression will be also be accessed in activated and resting primary T cells, to compare with the results obtained for Jurkat cells.

Maps of histone acetylation of the CD5 gene in T lymphocytes will be obtained by ChIP experiments, in order to confirm the higher levels in the pA1 PAS region compared with pA2 and pA3, as reported in other cell types in the UCSC Genome Browser.

The PTB IP experiment will be repeated, in order to improve the band visualization, while a SRSF6 IP will also be performed. In the meantime, overexpression of these factors will be performed, to further complement the knockdown results. In the long term, it would be interesting to unveil the mechanisms behind PTB and SRSF6 effects. As for SRSF3, it will be knocked-down in other genes to investigate if its effect is widespread, and overexpression will also be performed.

Bibliography

1. Kuby J., G.R.A., Kindt T. J. and Osborne B. A., *Immunology*. 5th ed. 2003.
2. Jones, N.H., et al., *Isolation of complementary DNA clones encoding the human lymphocyte glycoprotein T1/Leu-1.* Nature, 1986. **323**(6086): p. 346-9.
3. Gebe, J.A., et al., *Molecular cloning, mapping to human chromosome 1 q21-q23, and cell binding characteristics of Spalpha, a new member of the scavenger receptor cysteine-rich (SRCR) family of proteins.* J Biol Chem, 1997. **272**(10): p. 6151-8.
4. Calvo, J., et al., *Identification of a natural soluble form of human CD5.* Tissue Antigens, 1999. **54**(2): p. 128-37.
5. Azzam, H.S., et al., *CD5 Expression Is Developmentally Regulated By T Cell Receptor (TCR) Signals and TCR Avidity.* The Journal of Experimental Medicine, 1998. **188**(12): p. 2301-2311.
6. Dono, M., G. Cerruti, and S. Zupo, *The CD5+ B-cell.* Int J Biochem Cell Biol, 2004. **36**(11): p. 2105-11.
7. Youinou, P., C. Jamin, and P.M. Lydyard, *CD5 expression in human B-cell populations.* Immunology today, 1999. **20**(7): p. 312-316.
8. Van de Velde, H., et al., *The B-cell surface protein CD72/Lyb-2 is the ligand for CD5.* Nature, 1991. **351**(6328): p. 662-5.
9. Pospisil, R., et al., *CD5 is A potential selecting ligand for B-cell surface immunoglobulin: a possible role in maintenance and selective expansion of normal and malignant B cells.* Leuk Lymphoma, 2000. **36**(3-4): p. 353-65.
10. Biancone, L., et al., *Identification of a novel inducible cell-surface ligand of CD5 on activated lymphocytes.* J Exp Med, 1996. **184**(3): p. 811-9.
11. Brown, M.H. and E. Lacey, *A ligand for CD5 is CD5.* J Immunol, 2010. **185**(10): p. 6068-74.
12. Huang, H.J., et al., *Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells: molecular homology to its human counterpart Leu-1/T1 (CD5).* Proc Natl Acad Sci U S A, 1987. **84**(1): p. 204-8.
13. Vilà, J.M., et al., *Residues Y429 and Y463 of the human CD5 are targeted by protein tyrosine kinases.* European Journal of Immunology, 2001. **31**(4): p. 1191-1198.
14. Perez-Villar, J.J., et al., *CD5 negatively regulates the T-cell antigen receptor signal transduction pathway: involvement of SH2-containing phosphotyrosine phosphatase SHP-1.* Mol Cell Biol, 1999. **19**(4): p. 2903-12.
15. Calvo, J., et al., *Human CD5 Signaling and Constitutive Phosphorylation of C-Terminal Serine Residues by Casein Kinase II.* The Journal of Immunology, 1998. **161**(11): p. 6022-6029.
16. Sestero, C.M., et al., *CD5-Dependent CK2 Activation Pathway Regulates Threshold for T Cell Anergy.* The Journal of Immunology, 2012. **189**(6): p. 2918-2930.
17. Roa, N.S., et al., *The carboxy-terminal region of CD5 is required for c-CBL mediated TCR signaling downmodulation in thymocytes.* Biochem Biophys Res Commun, 2013. **432**(1): p. 52-9.
18. Dennehy, K.M., et al., *Thymocyte activation induces the association of the proto-oncoprotein c-cbl and ras GTPase-activating protein with CD5.* Eur J Immunol, 1998. **28**(5): p. 1617-25.
19. McAteer, M.J., et al., *A requirement for the CD5 antigen in T cell activation.* Eur J Immunol, 1988. **18**(7): p. 1111-7.
20. Ceuppens, J.L. and M.L. Baroja, *Monoclonal antibodies to the CD5 antigen can provide the necessary second signal for activation of isolated resting T cells by solid-phase-bound OKT3.* J Immunol, 1986. **137**(6): p. 1816-21.

21. Tarakhovsky, A., et al., *A role for CD5 in TCR-mediated signal transduction and thymocyte selection.* Science, 1995. **269**(5223): p. 535-7.
22. Pena-Rossi, C., et al., *Negative regulation of CD4 lineage development and responses by CD5.* J Immunol, 1999. **163**(12): p. 6494-501.
23. Raman, C., *CD5, an important regulator of lymphocyte selection and immune tolerance.* Immunol Res, 2002. **26**(1-3): p. 255-63.
24. Brossard, C., et al., *CD5 inhibits signaling at the immunological synapse without impairing its formation.* J Immunol, 2003. **170**(9): p. 4623-9.
25. Bamberger, M., et al., *A new pathway of CD5 glycoprotein-mediated T cell inhibition dependent on inhibitory phosphorylation of Fyn kinase.* J Biol Chem, 2011. **286**(35): p. 30324-36.
26. Azzam, H.S., et al., *Fine Tuning of TCR Signaling by CD5.* The Journal of Immunology, 2001. **166**(9): p. 5464-5472.
27. Fabbri, M., C. Smart, and R. Pardi, *T lymphocytes.* The International Journal of Biochemistry & Cell Biology, 2003. **35**(7): p. 1004-1008.
28. Friedlein, G., et al., *Human CD5 Protects Circulating Tumor Antigen-Specific CTL from Tumor-Mediated Activation-Induced Cell Death.* The Journal of Immunology, 2007. **178**(11): p. 6821-6827.
29. Dalloul, A., *CD5: A safeguard against autoimmunity and a shield for cancer cells.* Autoimmunity Reviews, 2009. **8**(4): p. 349-353.
30. Tabbekh, M., et al., *T-cell modulatory properties of CD5 and its role in antitumor immune responses.* OncoImmunology, 2013. **2**(1): p. e22841.
31. Bikah, G., et al., *CD5-mediated negative regulation of antigen receptor-induced growth signals in B-1 B cells.* Science, 1996. **274**(5294): p. 1906-9.
32. Lydyard, P.M., et al., *CD5+ B cells and the immune system.* Immunol Lett, 1993. **38**(2): p. 159-66.
33. Duan, B. and L. Morel, *Role of B-1a cells in autoimmunity.* Autoimmun Rev, 2006. **5**(6): p. 403-8.
34. Berland, R. and H.H. Wortis, *ORIGINS AND FUNCTIONS OF B-1 CELLS WITH NOTES ON THE ROLE OF CD5.* Annual Review of Immunology, 2002. **20**(1): p. 253-300.
35. Hawiger, D., et al., *Immunological unresponsiveness characterized by increased expression of CD5 on peripheral T cells induced by dendritic cells in vivo.* Immunity, 2004. **20**(6): p. 695-705.
36. Perez-Chacon, G., et al., *CD5 provides viability signals to B cells from a subset of B-CLL patients by a mechanism that involves PKC.* Leuk Res, 2007. **31**(2): p. 183-93.
37. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes Dev, 1997. **11**(21): p. 2755-66.
38. Wahle, E. and U. Ruegsegger, *3'-End processing of pre-mRNA in eukaryotes.* FEMS Microbiol Rev, 1999. **23**(3): p. 277-95.
39. Mohanty, B.K. and S.R. Kushner, *Bacterial/archaeal/organellar polyadenylation.* Wiley Interdiscip Rev RNA, 2011. **2**(2): p. 256-76.
40. Slomovic, S., et al., *Polyadenylation of ribosomal RNA in human cells.* Nucleic Acids Research, 2006. **34**(10): p. 2966-2975.
41. Bentley, D.L., *Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors.* Curr Opin Cell Biol, 2005. **17**(3): p. 251-6.
42. Zhao, J., L. Hyman, and C. Moore, *Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 405-45.
43. Slomovic, S., et al., *Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7407-12.

44. Beaudoing, E., et al., *Patterns of variant polyadenylation signal usage in human genes.* Genome Res, 2000. **10**(7): p. 1001-10.
45. Hirose, Y. and J.L. Manley, *RNA polymerase II is an essential mRNA polyadenylation factor.* Nature, 1998. **395**(6697): p. 93-6.
46. Proudfoot, N.J., *Ending the message: poly(A) signals then and now.* Genes Dev, 2011. **25**(17): p. 1770-82.
47. Shi, Y., *Alternative polyadenylation: new insights from global analyses.* Rna, 2012. **18**(12): p. 2105-17.
48. Tian, B. and J.L. Manley, *Alternative cleavage and polyadenylation: the long and short of it.* Trends Biochem Sci, 2013. **38**(6): p. 312-20.
49. Di Giammartino, Dafne C., K. Nishida, and James L. Manley, *Mechanisms and Consequences of Alternative Polyadenylation.* Molecular cell, 2011. **43**(6): p. 853-866.
50. Elkon, R., A.P. Ugalde, and R. Agami, *Alternative cleavage and polyadenylation: extent, regulation and function.* Nat Rev Genet, 2013. **14**(7): p. 496-506.
51. Lutz, C.S. and A. Moreira, *Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression.* Wiley Interdisciplinary Reviews: RNA, 2011. **2**(1): p. 22-31.
52. Zhang, H., J.Y. Lee, and B. Tian, *Biased alternative polyadenylation in human tissues.* Genome Biol, 2005. **6**(12): p. 28.
53. Sandberg, R., et al., *Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites.* Science, 2008. **320**(5883): p. 1643-7.
54. Ji, Z. and B. Tian, *Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types.* PLoS One, 2009. **4**(12): p. 0008419.
55. Bennett, C.L., et al., *A rare polyadenylation signal mutation of the FOXP3 gene (AAUAAA-->AAUGAA) leads to the IPEX syndrome.* Immunogenetics, 2001. **53**(6): p. 435-9.
56. Sanchez, G., et al., *Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer.* Proc Natl Acad Sci U S A, 2008. **105**(16): p. 6004-9.
57. Lu, F., A.B. Gladden, and J.A. Diehl, *An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene.* Cancer Res, 2003. **63**(21): p. 7056-61.
58. Wiestner, A., et al., *Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival.* Blood, 2007. **109**(11): p. 4599-606.
59. Mayr, C. and D.P. Bartel, *Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells.* Cell, 2009. **138**(4): p. 673-84.
60. Rehfeld, A., et al., *Alterations in polyadenylation and its implications for endocrine disease.* Front Endocrinol, 2013. **4**(53): p. 00053.
61. Wilusz, J., S.M. Pettine, and T. Shenk, *Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal.* Nucleic Acids Res, 1989. **17**(10): p. 3899-908.
62. Nunes, N.M., et al., *A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence.* EMBO J, 2010. **29**(9): p. 1523-36.
63. Ji, Z., et al., *Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 7028-33.
64. Takagaki, Y., et al., *The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation.* Cell, 1996. **87**(5): p. 941-52.

65. Bruce, S.R., R.W. Dingle, and M.L. Peterson, *B-cell and plasma-cell splicing differences: a potential role in regulated immunoglobulin RNA processing.* Rna, 2003. **9**(10): p. 1264-73.
66. Elkon, R., et al., *E2F mediates enhanced alternative polyadenylation in proliferation.* Genome Biol, 2012. **13**(7).
67. Martin, G., et al., *Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length.* Cell Rep, 2012. **1**(6): p. 753-63.
68. Spies, N., et al., *Biased chromatin signatures around polyadenylation sites and exons.* Mol Cell, 2009. **36**(2): p. 245-54.
69. Khaladkar, M., M. Smyda, and S. Hannenhalli, *Epigenomic and RNA structural correlates of polyadenylation.* RNA Biol, 2011. **8**(3): p. 529-37.
70. Lee, C.Y. and L. Chen, *Alternative polyadenylation sites reveal distinct chromatin accessibility and histone modification in human cell lines.* Bioinformatics, 2013. **29**(14): p. 1713-7.
71. Cowley, M., et al., *Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus.* Nucleic Acids Res, 2012. **40**(18): p. 8917-26.
72. Sims, R.J. and D. Reinberg, *Processing the H3K36me3 signature*. 2009: Nat Genet. 2009 Mar;41(3):270-1. doi: 10.1038/ng0309-270.
73. Pandit, S., D. Wang, and X.D. Fu, *Functional integration of transcriptional and RNA processing machineries.* Curr Opin Cell Biol, 2008. **20**(3): p. 260-5.
74. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription.* Cell, 2002. **108**(4): p. 501-12.
75. Mandel, C.R., Y. Bai, and L. Tong, *Protein factors in pre-mRNA 3'-end processing.* Cell Mol Life Sci, 2008. **65**(7-8): p. 1099-122.
76. Dantonel, J.C., et al., *Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA.* Nature, 1997. **389**(6649): p. 399-402.
77. Barilla, D., B.A. Lee, and N.J. Proudfoot, *Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2001. **98**(2): p. 445-50.
78. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription.* Nature, 1997. **385**(6614): p. 357-61.
79. Nagaike, T., et al., *Transcriptional activators enhance polyadenylation of mRNA precursors.* Mol Cell, 2011. **41**(4): p. 409-18.
80. Pinto, P.A., et al., *RNA polymerase II kinetics in polo polyadenylation signal selection.* Embo J, 2011. **30**(12): p. 2431-44.
81. Moreira, A., *Integrating transcription kinetics with alternative polyadenylation and cell cycle control.* Nucleus, 2011. **2**(6): p. 556-61.
82. Kaida, D., et al., *U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation.* Nature, 2010. **468**(7324): p. 664-8.
83. Gunderson, S.I., M. Polycarpou-Schwarz, and I.W. Mattaj, *U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase.* Mol Cell, 1998. **1**(2): p. 255-64.
84. Hall-Pogar, T., et al., *Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR.* Rna, 2007. **13**(7): p. 1103-15.
85. Berg, M.G., et al., *U1 snRNP determines mRNA length and regulates isoform expression.* Cell, 2012. **150**(1): p. 53-64.
86. Millevoi, S. and S. Vagner, *Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation.* Nucleic Acids Res, 2010. **38**(9): p. 2757-74.
87. Licatalosi, D.D., et al., *HITS-CLIP yields genome-wide insights into brain alternative RNA processing.* Nature, 2008. **456**(7221): p. 464-9.

88. Calado, A., et al., *Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA.* Hum Mol Genet, 2000. **9**(15): p. 2321-8.
89. Jenal, M., et al., *The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites.* Cell, 2012. **149**(3): p. 538-53.
90. Hilgers, V., S.B. Lemke, and M. Levine, *ELAV mediates 3' UTR extension in the Drosophila nervous system.* Genes Dev, 2012. **26**(20): p. 2259-64.
91. Moreira, A., et al., *The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms.* Genes Dev, 1998. **12**(16): p. 2522-34.
92. Castelo-Branco, P., et al., *Polypyrimidine tract binding protein modulates efficiency of polyadenylation.* Mol Cell Biol, 2004. **24**(10): p. 4174-83.
93. Danckwardt, S., et al., *Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals.* Embo J, 2007. **26**(11): p. 2658-69.
94. Millevoi, S., et al., *A physical and functional link between splicing factors promotes pre-mRNA 3' end processing.* Nucleic Acids Res, 2009. **37**(14): p. 4672-83.
95. Costessi, L., et al., *Characterization of the distal polyadenylation site of the ss-adducin (Add2) pre-mRNA.* PLoS One, 2013. **8**(3): p. e58879.
96. Blechingberg, J., et al., *Regulatory mechanisms for 3'-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript.* Nucleic Acids Res, 2007. **35**(22): p. 7636-50.
97. Maciolek, N.L. and M.T. McNally, *Serine/arginine-rich proteins contribute to negative regulator of splicing element-stimulated polyadenylation in rous sarcoma virus.* J Virol, 2007. **81**(20): p. 11208-17.
98. Dettwiler, S., et al., *Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization.* J Biol Chem, 2004. **279**(34): p. 35788-97.
99. Cui, M., et al., *Genes involved in pre-mRNA 3'-end formation and transcription termination revealed by a lin-15 operon Muv suppressor screen.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16665-70.
100. Ciafre, S.A. and S. Galardi, *microRNAs and RNA-binding proteins: A complex network of interactions and reciprocal regulations in cancer.* RNA Biol, 2013. **10**: p. 6.
101. Kedde, M. and R. Agami, *Interplay between microRNAs and RNA-binding proteins determines developmental processes.* Cell Cycle, 2008. **7**(7): p. 899-903.
102. Bas, A., et al., *Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes.* Scand J Immunol, 2004. **59**(6): p. 566-73.
103. Aerts, J.L., M.I. Gonzales, and S.L. Topalian, *Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR.* Biotechniques, 2004. **36**(1): p. 84-6.
104. Domingues, R.G., *CD5 Alternative Polyadenylation and T Lymphocyte Activation.* Master's thesis, 2010.
105. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-9.
106. Biselli, R., et al., *Multiparametric flow cytometric analysis of the kinetics of surface molecule expression after polyclonal activation of human peripheral blood T lymphocytes.* Scand J Immunol, 1992. **35**(4): p. 439-47.

107. Lim, L.C., et al., *A whole-blood assay for qualitative and semiquantitative measurements of CD69 surface expression on CD4 and CD8 T lymphocytes using flow cytometry.* Clin Diagn Lab Immunol, 1998. **5**(3): p. 392-8.
108. Akerman, M., et al., *A computational approach for genome-wide mapping of splicing factor binding sites.* Genome Biol, 2009. **10**(3): p. 2009-10.
109. Carthew, R.W. and E.J. Sontheimer, *Origins and Mechanisms of miRNAs and siRNAs.* Cell, 2009. **136**(4): p. 642-55.
110. Gavrilov, K. and W.M. Saltzman, *Therapeutic siRNA: principles, challenges, and strategies.* Yale J Biol Med, 2012. **85**(2): p. 187-200.
111. Gooding, C., et al., *MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3.* Nucleic Acids Res, 2013. **41**(9): p. 4765-82.
112. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it.* Genes Dev, 2004. **18**(20): p. 2437-68.
113. Kurdistani, S.K., S. Tavazoie, and M. Grunstein, *Mapping global histone acetylation patterns to gene expression.* Cell, 2004. **117**(6): p. 721-33.